

UNIVERSIDADE FEDERAL DO PARANÁ

ELIZEU MARTINS DE OLIVEIRA JUNIOR

**PROPOSTA DE UM MÉTODO ADAPTATIVO PARA REGISTRO DE
DADOS RGB-D**

**CURITIBA
2015**

ELIZEU MARTINS DE OLIVEIRA JUNIOR

**PROPOSTA DE UM MÉTODO ADAPTATIVO PARA REGISTRO DE
DADOS RGB-D**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Geodésicas da Universidade Federal do Paraná, como requisito final para cumprimento dos créditos em disciplina.

**Orientador: Prof. Dr. Daniel R. dos Santos
Coorientador: Prof. Dr. Quintino Dalmolin**

**CURITIBA
2015**

ELIZEU MARTINS DE OLIVEIRA JUNIOR

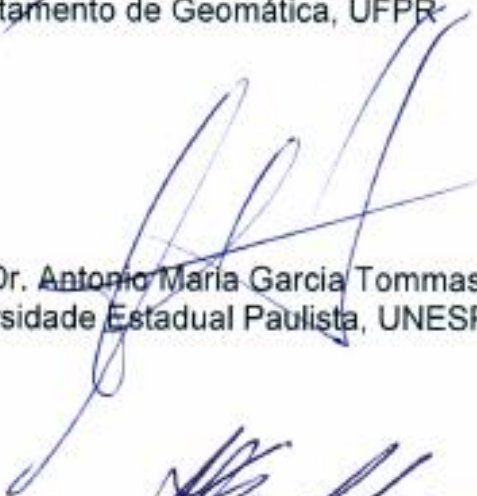
"PROPOSTA DE UM MÉTODO ADAPTATIVO
PARA REGISTRO DE DADOS RGB-D"

Dissertação nº 292 aprovada como requisito parcial do grau de Mestre no Programa de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador:



Prof. Dr. Daniel Rodrigues dos Santos
Departamento de Geomática, UFPR



Prof. Dr. Antonio Maria Garcia Tommaselli
Universidade Estadual Paulista, UNESP



Prof. Dr. Jorge Antonio Silva Centeno
Departamento de Geomática, UFPR

Curitiba, 04 de setembro de 2015.

DEDICATÓRIA

Dedico este trabalho à minha esposa Hellen, que sempre me apoiou no meu crescimento profissional, mesmo quando essa busca implicou na mudança de cidade e aos meus pais Jocirlei e Elizeu.

AGRADECIMENTOS

A Deus.

A minha esposa e a toda minha família.

Ao Professor Dr. Daniel Rodrigues dos Santos, pela orientação, confiança, apoio, compreensão, conselhos e amizade.

Aos membros das bancas do Seminário I e de defesa, Professor Dr. Antonio Maria Garcia Tommaselli, Professor Dr. Jorge Centeno e Professor Dr. Quintino Dalmolin pelas sugestões oferecidas.

A todos os professores do Programa de Pós-Graduação em Ciências Geodésicas em especial aos Professores Dr. Álvares Machado, Dr. Silvio Freitas, Dr. Edson Aparecido Mitishita, Dr^a. Cláudia Robbi Sluter e Dr^a. Luciene S. Delazari.

Aos meus colegas e amigos Gabriel Dresch, Nadisson Luis Pavan, Giovana Ros Miola, Jaqueline Prunzel, Andrey, André, Gabriela, Letícia, Adriana, Leonardo, Gustavo, Wander, Nícolas, Kauê, Marcos, Niarkos e João.

Aos Professores da UNEMAT em especial aos professores Érico Fernando de Oliveira Martins e Rodrigo Bruno Zanin pelo apoio e incentivo.

A Mônica, secretária do Programa de Pós-Graduação em Ciências Geodésicas.

Agradeço à Universidade Federal do Paraná pela oportunidade de aperfeiçoar a minha formação profissional e a CAPES pelo fornecimento da bolsa de estudos.

RESUMO

Neste trabalho é apresentado um método adaptativo para mapeamento 3D de ambientes internos usando dados derivados de câmeras RGB-D. Basicamente, quatro pontos principais envolvidos no desenvolvimento do método são aqui discutidos: a calibração dos sensores integrados no dispositivo Kinect; o registro dos pares de nuvens de pontos 3D adquiridas sucessivamente pelo sensor de imageamento; a inclusão de pesos no algoritmo SIFT baseados na precisão teórica das medidas de profundidade; e o refinamento dos parâmetros de translação empregando uma análise de consistência global. A calibração dos sensores é feita usando o método de Zhang. O modelo de corpo rígido 3D é empregado para calcular os parâmetros de transformação com uma correspondência ponto-ponto. Um modelo linear capaz de refinar simultaneamente o conjunto de parâmetros de translação é proposto para a análise da consistência global. Experimentos foram conduzidos para avaliar a eficiência do método proposto e os resultados obtidos mostraram uma melhoria significativa no registro do conjunto de nuvens de pontos 3D com ponderação na coordenada Z .

Palavras chaves: Câmeras RGB-D; Calibração de câmeras; Mapeamento de ambientes internos; Registro de pares de nuvens de pontos; Análise de consistência global.

ABSTRACT

In this paper is presented an adaptive method for 3D indoor mapping using RGB-D data. Basically, four main points involved in the method are here discussed: the calibration of the sensors integrated in the Kinect device; the registration of 3D point clouds pairs acquired successively by the imaging sensor; the inclusion of weights in SIFT algorithms based in theoretical random error of depth measurements; and a global adjustment for the refinement of translation parameters. The sensors calibration is realized using the Zhang method. The 3D rigid body model is employed to calculate the transformation parameters with based point-point correspondence. A linear model is proposed for global adjustment. Experiments were conducted to evaluate the efficiency of the proposed method and the results obtained show a significant improvement in 3D cloud points set registering with weight in depth coordinate.

Keywords: RGB-D cameras; Camera calibration; Indoor mapping; Registration of RGB-D data; Global adjustment.

LISTA DE ILUSTRAÇÕES

FIGURA 1 - SENSORES INTEGRANTES DO KINECT	23
FIGURA 2 - PADRÃO DE LUZ ESTRUTURADA GERADO PELO KINECT	24
FIGURA 3 - ANALOGIA ENTRE PROFUNDIDADE RELATIVA E A PARALAXE MEDIDA	25
FIGURA 4 - REPRESENTAÇÃO DO PROCESSO PARA CONSTRUÇÃO DAS IMAGENS RESULTANTES DA DIFERENÇA DE GAUSSIANS EM CADA OITAVA NO ESPAÇO ESCALA.....	32
FIGURA 5 - EXEMPLO DO PROCESSO DE DETECÇÃO DE EXTREMOS NO ESPAÇO-ESCALA.....	33
FIGURA 6 - HISTOGRAMA DE ORIENTAÇÕES DE UM PONTO VISUAL.....	35
FIGURA 7 - CONSTRUÇÃO DO DESCRITOR PARA UM PONTO VISUAL	36
FIGURA 8 – PROCESSO DE ESTABELECIMENTO DE PSEUDO-CONJUGADOS ENTRE DUAS NUVENS QUE REPRESENTAM A MESMA SUPERFÍCIE PELO MÉTODO DE VIZINHO MAIS PRÓXIMO.....	42
FIGURA 9 - POSES DESALINHADAS DEVIDO AO ACÚMULO DE ERROS	44
FIGURA 10 - DIAGRAMA DO MÉTODO PROPOSTO	47
FIGURA 11- EXPERIMENTO REALIZADO COM O INTERFERÔMETRO E COM O KINECT PARA A NORMALIZAÇÃO DOS VALORES DE PROFUNDIDADE. (a) IMAGEM RGB; (b) IMAGEM DE PROFUNDIDADE.....	50
FIGURA 12 – SEQUÊNCIA DE PASSOS PARA A ETAPA DO REGISTRO GROSSEIRO.....	52
FIGURA 13- PROCESSO DE REMOÇÃO DE OUTLIERS COM O RANSAC	55
FIGURA 14 - SEQUÊNCIA DE PASSOS PARA A ETAPA DO REFINAMENTO DOS PARÂMETROS	61
FIGURA 15 - CONJUNTO DE IMAGENS RGB E IR OBTIDAS PARA CALIBRAÇÃO DOS SENSORES. (A) IMAGENS IR; (B) IMAGENS RGB.....	66
FIGURA 16 – RELAÇÃO LINEAR DA PARALAXE NORMALIZADA COM O INVERSO DA PROFUNDIDADE.....	68
FIGURA 17 - TRAJETÓRIA DO SENSOR DEPOIS DA ANÁLISE DA CONSISTÊNCIA GLOBAL E O VETOR DAS DISCREPÂNCIAS MÉDIAS. (a) CENÁRIO A; (b) CENÁRIO B; (c) CENÁRIO C.....	70
FIGURA 18 - TRAJETÓRIA DO SENSOR NO CENÁRIO D DEPOIS DA ANÁLISE DE CONSISTÊNCIA GLOBAL E ANTES COM PESO EM Z (a), EM XYZ (b) E SEM PESOS (c).....	72

FIGURA 19 - TRAJETÓRIA DO SENSOR NO CENÁRIO E DEPOIS DA ANÁLISE DE CONSISTÊNCIA GLOBAL E ANTES COM PESO EM Z (a), EM XYZ (b) E SEM PESOS (c).....	73
FIGURA 20 - TRAJETÓRIA DO SENSOR NO CENÁRIO J DEPOIS DA ANÁLISE DE CONSISTÊNCIA GLOBAL E ANTES COM PESO EM Z (a), EM XYZ (b) E SEM PESOS (c).....	75
FIGURA 21 - NUVEM DE PONTOS 3D DE AMBIENTES INTERNOS MAPEADOS. (a) CENÁRIO A; (b) CENÁRIO B; (c) CENÁRIO C	76
FIGURA 22 - NUVEM DE PONTOS 3D DOS AMBIENTES INTERNOS. (a) CENÁRIO F; (b) CENÁRIO G; (c) CENÁRIO H	78
FIGURA 23 - REMQ DA MÉDIA DAS DISCREPÂNCIAS CALCULADA EM CADA REGISTRO DE PARES DE NUVENS DE PONTOS. (a) CENÁRIO D; (b) CENÁRIO I; (c) CENÁRIO E.	80
FIGURA 24 - INLIERS ADVINDOS DO REGISTRO GROSSEIRO COM O MÉTODO SIFT. (a) CENÁRIO I; (b) CENÁRIO J	81
FIGURA 25 - DISTÂNCIA (AZUL) E ÂNGULO (VERMELHO) DE FECHAMENTO PARA SEQUÊNCIA DE REGISTROS ENTRE PARES DE NUVENS DE PONTOS COM E SEM PESO. (a) CENÁRIO D; (b) CENÁRIO E;.....	83
FIGURA 26 - DISTRIBUIÇÃO DOS VALORES DE PESO PARA AS OBSERVAÇÕES. (a) INFLUÊNCIA DOS VALORES DE PESOS EM X; (b) INFLUÊNCIA DOS VALORES DE PESOS EM Y; (c) INFLUÊNCIA DOS VALORES DE PESOS EM Z.	85

LISTA DE TABELAS

TABELA 1 - PARÂMETROS DE ORIENTAÇÃO INTERIOR DOS SENSORES RGB E IR, SEUS DESVIOS PADRÃO E OS VALORES DOS POE.....	67
TABELA 2 - MÉDIA DAS DISCREPÂNCIAS APÓS A ANÁLISE DA CONSISTÊNCIA GLOBAL PARA OS CENÁRIOS A, B e C	80
TABELA 3 - VALORES MÁXIMOS E MÍNIMOS DA REMQ PARA OS CENÁRIOS F, G e H.....	80

LISTA DE QUADROS

QUADRO 1 - DETERMINAÇÃO DOS PARÂMETROS DE ORIENTAÇÃO INTERIOR	29
QUADRO 2 - EXPERIMENTOS REALIZADOS	68

LISTA DE SIGLAS

3D - Tridimensional

BBF - Best-Bin-First

CMOS - *Complementary Metal-Oxide-Semiconductor*

DoG - *Difference of Gaussians*

fps – *Frames/quadros por segundo*

G²O - *General framework for graph Optimization*

ICP - *Iterative Closest Point*

IMU - *Inertial Measurement Unit*

IR - *InfraRed*

K-D Tree - K-Dimensional Tree

LAPE - Laboratório de Aquisição e Processamento de Dados Espaciais

LASER - *Light Amplification by Stimulated Emission of Radiation*

MF - Matriz Fundamental

MMQ - Método dos Mínimos Quadrados

MRPT - *Mobile Robot Programming Toolkit*

MVC - Matriz Variância Covariância

OpenCV - *Open-Source Computer Vision*

POE – Parâmetros de Orientação Exterior

POI – Parâmetros de Orientação Interior

RANSAC - *Randon Sample Consensus*

REMQ - Raiz do Erro Médio Quadrático

RGB - *Red Green e Blue*

RGB-D - *Red Green e Blue + Depth*

SBA - *Sparse Bundle Adjustment*

SMAI-3D - Sistemas de Mapeamento 3D de Ambientes Internos

SIFT - *Scale Invariant Feature Transform*

SVD - *Singular Value Decomposition*

TORO - *Tree-Based netwORk Optimizer*

SUMÁRIO

1.	INTRODUÇÃO	14
1.1.	CONSIDERAÇÕES INICIAIS	14
1.2.	OBJETIVO GERAL	16
1.3.	OBJETIVOS ESPECÍFICOS	16
1.4.	JUSTIFICATIVA E CONTRIBUIÇÕES	17
1.5.	ESTRUTURA E CONTEÚDO DO TRABALHO	17
2.	FUNDAMENTAÇÃO TEÓRICA	18
2.1.	ESTADO DA ARTE	18
2.2.	DISPOSITIVO KINECT	23
2.2.1.	Calibração de Sensores	27
2.3.	REGISTRO DOS PARES DE NUVENS DE PONTOS 3D	30
2.3.1.	Determinação dos parâmetros de transformação do registro grosseiro	30
2.3.4.	Algoritmo ICP	41
2.4.	AJUSTAMENTO SEQUENCIAL DE PARES DE NUVENS DE PONTOS	43
3.	MATERIAIS E MÉTODOS	45
3.3.	MATERIAIS	45
3.3.1.	Laboratório e Recursos Bibliográficos	45
3.3.2.	Recursos de Hardware	45
3.3.3.	Recursos de Software e Bibliotecas Computacionais <i>Open-Source</i>	46
3.4.	MÉTODO	46
3.4.1.	Calibração do sistema	48
3.4.2.	Registro de Pares de Nuvens de Pontos	50
3.4.3.	Ajustamento por pares de nuvens de pontos independentes	64
4.	EXPERIMENTOS E RESULTADOS	66
5.	CONCLUSÕES E RECOMENDAÇÕES	86
	REFERÊNCIAS	89

1. INTRODUÇÃO

1.1. CONSIDERAÇÕES INICIAIS

De acordo com o relatório de 2010 da Organização Mundial da Saúde (OMS) metade da população mundial reside em áreas urbanizadas. Embora, desde 2400 a.c., o mapa tem sido empregado como uma ferramenta crucial para auxiliar na localização e na navegação, apenas no final do século XX foram desenvolvidos os primeiros Sistemas de Mapeamento 3D de Ambientes Internos (SMAI-3D). Um sistema como este possui diversas vantagens e pode ser empregado em inúmeras aplicações, tais como, em problemas de localização e mapeamento simultâneo (SLAM), mapeamento de corredores, de salas, de túneis e de cavernas, modelagem e reconstrução 3D de superfícies, navegação, vigilância e aplicações forenses, realidade aumentada, estabelecimento de rotas de fuga em edifícios em caso de incêndios ou outros incidentes, entre outras.

Os primeiros SMAI-3D foram desenvolvidos baseados em sensores de imageamento ativo (varredura LASER-*Light Amplification by Stimulated Emission of Radiation* e câmeras de distância 3D) e passivos (mono ou estéreo câmeras) embarcados em plataformas fixas ou móveis. Usualmente, um SMAI-3D é dividido em três partes principais: o registro dos dados tridimensionais (nuvem de pontos 3D) adquiridos pelo sensor de imageamento em diferentes pontos de vista; a detecção de lugares anteriormente visitados; e, o refinamento da trajetória estimada (análise de consistência global). Das três, a mais importante é a do registro dos dados, que consiste basicamente na determinação dos parâmetros de rotação e translação relativa entre os pares de nuvens de pontos consecutivos, por intermédio de um modelo matemático que descreva os parâmetros mencionados e um método de ajustamento de observações. Em sua maioria, os métodos propostos para o registro de nuvens de pontos são divididos em analíticos (FAUGERAS e HELBERT, 1986; HORN, 1987; ARUN *et al.*, 1987) e iterativos (BESL; MCKAY, 1992; CHEN e MEDIONI, 1992; ZHANG, 1994, BLAIS e LEVINE, 1995; MASUDA *et al.*, 1996; RUSISKIEWICZ e LEVOY, 2001; GRUEN e AKCA, 2005; NUCHTER *et al.*, 2007; SEGAL *et al.*, 2009; SANDE *et al.*, 2010; GRANT *et al.*, 2012). O método iterativo

mais utilizado pela comunidade científica é o ICP (*Iterative Closest Point*), originalmente proposto por Besl e McKay (1992). Basicamente, o algoritmo encontra pontos pseudo-conjugados e, iterativamente, determina os parâmetros do registro (3 rotações e 3 translações) entre os pares de nuvens de pontos 3D. Isto é feito por intermédio de um modelo matemático de transformação rígida e uma função que minimiza a distância entre os pares de primitivas correspondentes. No entanto, a eficiência do ICP depende da entrada de valores iniciais aproximados e o processamento dos dados é computacionalmente lento, caso não haja alta sobreposição entre as nuvens do par.

Para contornar este problema Magee *et al.* (1985), Godin *et al.* (1994), Weik (1997), Roth (1999), Stamos e Leordeau (2003), Bendels *et al.* (2004), Al-manasir e Fraser (2006), Ellekilde *et al.* (2007), Huhle *et al.* (2008), Barnea e Filin (2008) e Mostofi *et al.* (2014), exploraram as vantagens sinérgicas advindas da ação simultânea das informações derivadas de imagens RGB e de profundidade. A ideia principal é dividir a tarefa em duas etapas: primeiro, são obtidos os parâmetros do registro grosseiro entre pares de nuvens de pontos; e, em seguida, é feito um refinamento desses parâmetros através do algoritmo ICP melhorando seu desempenho e eficiência computacional. Vale ressaltar que a sinergia supracitada dependia da integração de sensores de varredura LASER e ópticos de alta resolução.

Atualmente, câmeras RGB-D (Kinect) têm sido amplamente empregadas em problemas de mapeamento 3D de ambientes internos, por serem dispositivos de baixo custo (valor de venda gira em torno de R\$350,00), leves e de fácil manuseio, além de fornecerem uma sinergia natural de dados derivados de sensores RGB e de profundidade. Comparado com o sensor LASER *scanning* terrestre e com uma câmera de distância 3D, o Kinect apresenta pior precisão e resolução geométrica (KHOSHELHAM e ELBERINK, 2012), entretanto, além de ser mais barato, sua alta taxa de aquisição de dados tornam o sensor um dispositivo atraente para mapeamento 3D de ambientes internos. Apesar de serem originalmente construídos para finalidades de entretenimento e jogos computacionais, suas características e facilidades de aquisição de dados, bem como o tipo de informação fornecida chamou a atenção de pesquisadores em diversas áreas do conhecimento, como por exemplo, na Fotogrametria, na Visão Computacional e na Robótica.

O primeiro SMAI-3D usando câmeras RGB-D foi proposto por Henry *et al.* (2012). Entretanto, neste trabalho a associação entre os pontos correspondentes na RGB e os valores de profundidade é feita de forma aproximada, ou seja, o desalinhamento entre os sensores RGB e IR é negligenciado. Os autores também não admitiram, no processo de registro dos pares de nuvens de pontos, as incertezas provenientes das coordenadas tridimensionais. O resultado esperado é uma trajetória do sensor de baixa qualidade.

Visando solucionar os problemas apontados, neste trabalho é proposta a inclusão de pesos no processo de registro dos pares de nuvens de pontos 3D para obtenção de uma trajetória do sensor mais robusta e de melhor acurácia relativa. Também, é proposto um modelo linear para o ajustamento simultâneo em bloco do conjunto de pares de nuvens de pontos capaz de refinar os parâmetros de translação do registro baseado no método de LUM (LU e MILLIOS, 1997).

1.2. OBJETIVO GERAL

O objetivo geral deste trabalho é investigar e desenvolver um método adaptativo para registro de dados RGB-D.

1.3. OBJETIVOS ESPECÍFICOS

Os objetivos específicos a serem atingidos são os seguintes:

- a) Realizar a calibração dos sensores RGB e IR, bem como determinar os parâmetros de montagem do dispositivo Kinect;
- b) Desenvolver uma abordagem adaptativa para a etapa de registro dos pares de nuvens de pontos 3D com inclusão de pesos baseado na precisão teórica das coordenadas X , Y e Z ;
- c) Mostrar as influências exercidas pelas precisões teóricas das coordenadas X , Y e Z na trajetória do sensor, durante o processo de mapeamento;

d) Desenvolver um modelo linear para ajustamento simultâneo dos pares de nuvens de pontos 3D tratando o problema de consistência global em modo pós-processamento a partir da Matriz Variância Covariância (MVC) dos parâmetros de registro;

1.4. JUSTIFICATIVA E CONTRIBUIÇÕES

As principais razões que justificam o desenvolvimento deste projeto são: a demanda por sistemas de mapeamento 3D em ambientes internos e a redução do esforço operacional na sistematização do mapeamento através de técnicas fotogramétricas; o desenvolvimento e aquisição do conhecimento da nova tecnologia para fins de monitoramento e mapeamento 3D em ambientes internos; e a relevância científica comprovada nas últimas publicações encontradas nos principais periódicos da área e afins.

A principal contribuição do método proposto é a concepção de uma abordagem adaptativa para a etapa de registro dos pares de nuvens de pontos 3D, tendo como característica a associação rigorosa dos pares de pontos correspondentes com a nuvem de pontos baseada na retificação da imagem IR em relação a imagem RGB, a inclusão de pesos em função da precisão teórica das coordenadas X , Y e Z , e o desenvolvimento de um modelo linear para a etapa do ajustamento simultâneo dos pares de nuvens de pontos.

1.5. ESTRUTURA E CONTEÚDO DO TRABALHO

A estrutura do trabalho está subdividida em quatro capítulos. No Capítulo 2 será apresentado o estado da arte dos trabalhos que fundamentam o Capítulo de Materiais e Métodos, e uma revisão bibliográfica dos principais conceitos utilizados. No Capítulo 3 serão descritos os Materiais e o Método proposto. Os experimentos e a discussão dos resultados serão apresentados no Capítulo 4. Finalmente, as conclusões e recomendações para trabalhos futuros serão apresentadas no Capítulo 5.

2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo apresentar a fundamentação teórica do mapeamento de ambientes internos usando sensores RGB-D. Primeiramente será apresentado o estado da arte, em seguida, o princípio de funcionamento do dispositivo Kinect, o modelo matemático básico para a determinação da nuvem de pontos 3D e uma técnica de calibração dos sensores RGB-D baseado no método de Zhang (2000). Posteriormente, será apresentado o algoritmo SIFT para detecção de pontos visuais na imagem RGB e uma técnica robusta para a associação desses pontos à nuvem de pontos 3D. Finalmente, é apresentado o método paramétrico de ajustamento pelo Método dos Mínimos Quadrados (MMQ) para estimar os parâmetros do registro grosseiro.

2.1. ESTADO DA ARTE

Devido ao enorme sucesso das câmeras RGB-D, sistemas baseados em sensores ópticos tem se tornado uma das opções mais econômicas, rápidas e seguras embarcadas em SMAI-3D. Basicamente, esses dispositivos propiciam uma nuvem de pontos 3D com precisão de 3 centímetros a cada 3 metros e são capazes de fornecer em torno de 9,2 milhões de pontos por segundo, valor muito acima de qualquer outro sistema LASER scanning terrestre. Em Khoshelham e Elberink (2012), Menna *et al.* (2011) e Chow *et al.* (2012), a resolução geométrica e a precisão dos pontos 3D oferecidos pelo dispositivo são tratados como objeto de estudo. Esses trabalhos mostraram a importância na calibração dos sensores dispostos no Kinect. Os dois primeiros são baseados na calibração de câmeras com uso de valores de paralaxe fornecidos pelo dispositivo Kinect. Já no trabalho de Chow *et al.* (2012), uma nova rotina de calibração é apresentada para a modelagem da câmera IR, RGB e do projetor (tratado como uma câmera reversa) usando a equação de colinearidade com *boresight*, *lever arm*, coordenadas no espaço objeto dos alvos medidos pela câmera RGB e os parâmetros do plano (as primitivas de seu modelo funcional são ponto-plano) simultaneamente em um modelo combinado de

ajustamento por Mínimos Quadrados. Entretanto, os Parâmetros de Orientação Interior (POI), Exterior (POE) e adicionais (coeficientes de distorção das lentes e de assimetria) da câmera IR, e, do projetor não puderam ser recuperados. Todavia, em Chow e Lichti (2013) este problema é resolvido iterativamente através de projeções e reprojeções dada a natureza não-linear das equações de colinearidade. Chow e Lichti (2013) classificam seu método como: preciso, uma vez que as coordenadas do alvo no espaço objeto, parâmetros do plano, POI, POE e adicionais, dos três sensores, são estimados simultaneamente no mesmo ajuste por feixes de raios; prático, já que apenas um alvo simples com o desenho de um tabuleiro de xadrez foi usado; amplamente aplicável, logo que são usadas medidas de profundidade ao invés de descontinuidades de profundidade encontradas nos limites do plano.

Henry *et al.* (2010), Du *et al.* (2011), Engelhard *et al.* (2011), Bachrach *et al.* (2012), Dryanovsky *et al.* (2012), Endres *et al.* (2012), Henry *et al.* (2012), entre outros, propuseram métodos para mapeamento 3D de ambientes internos baseado em abordagens *coarse-to-fine*. Primeiramente, primitivas pontuais detectadas em imagens RGB são empregadas para determinar os parâmetros de transformação do registro grosseiro entre pares de nuvens de pontos 3D seguido de um refinamento dos dados usando o algoritmo ICP. Já para o ajustamento por pares de nuvens de pontos independentes, os autores empregaram métodos baseados em abordagens probabilísticas, tais como: o método de otimização baseado em árvore de busca (GRISSETTI *et al.* 2007, *Tree-based netwORk Optimizer* - TORO); o método de otimização baseado em grafos (KÜMMERLE *et al.* 2011, *General framework for graph Optimization* – G²O); e o método que considera as relações espaciais entre as nuvens como variáveis randômicas (LU e MILIOS 1997 - LUM). Também vale citar o método de ajustamento por feixes perspectivos (*Sparse Bundle Adjustment* - SBA) proposto por Lourakis e Argyros (2009), o *GraphSLAM*, e o *Occupancy Grid* (THRUN *et al.*, 2009).

Em Henry *et al.* (2010), é apresentado o *RGB-D Mapping*, que explora tanto as feições na nuvem de pontos como informações na imagem RGB para o estabelecimento de correspondências entre pontos e, assim, fazer o registro das nuvens. Nessa abordagem, erros de fechamento são detectados através de processos de correlação com elementos estruturantes previamente armazenadas pelo algoritmo. Para registrar todas as nuvens a um único sistema global é utilizado o método TORO. Segundo os autores, uma correta integração entre as medidas na

imagem RGB às de profundidade proporciona mais robustez ao método de registro dos pares de nuvens de pontos e de detecção de erros de fechamento.

Du *et al.* (2011) desenvolveram um sistema interativo de mapeamento 3D de ambientes internos voltado ao público em geral. É um sistema *on-line* que transmite as nuvens de pontos 3D coloridas reconstruídas simultaneamente para um *laptop*. Para tanto, informações da imagem RGB são utilizadas para auxiliar o processo de registro dos pares de nuvens de pontos. Para este sistema, o usuário é advertido quando diante de cenas pouco texturizadas (que possui igual natureza e/ou apresenta semelhança de estrutura), para que retorne a determinada posição permitindo ao algoritmo recalcular aquela posição. Essa interação também acontece na parte de detecção de erros de fechamento, dessa forma, o usuário verifica e adiciona links entre feições em diferentes imagens facilitando na tomada de decisão do algoritmo. O método de registro usado por esse sistema *on-line* combina RANSAC (*Randon Sample Consensus*) com conflitos de 'visibilidade'. Apesar da atratividade, o método peca em dois principais aspectos: Com a participação humana, erros grosseiros podem ser inseridos no processo; O segundo aspecto reside em assumir que as imagens de profundidade e RGB estejam devidamente alinhadas.

Em Engelhard *et al.* (2011) é realizado o registro das nuvens de pontos por meio de primitivas pontuais detectadas nas imagens RGB, associadas às nuvens de pontos, e, portanto, estimados os parâmetros por RANSAC. Em seguida, é realizado um refinamento desses parâmetros com a técnica ICP e o grafo das posições é otimizado com a técnica TORO. Entretanto, além de não haver nenhum controle à associação dos pontos da imagem RGB às nuvens de pontos 3D, não é realizado nenhum ajustamento desses parâmetros, já que somente com o RANSAC esse registro apresenta uma qualidade ainda mais inferior quando comparado ao resultado refinado, o que pode onerar ainda mais o algoritmo ICP.

Em Bachrach *et al.* (2012) é realizada a integração de um sensor RGB-D a um DRONE sendo que o método para localização e mapeamento do sistema é também responsável por orientar a trajetória do dispositivo. Feições visuais são detectadas nas imagens RGB, em seguida, são associadas às imagens de profundidade e a rotação é estimada por correlação de imagens. Descritores visuais são usados para atribuir as correspondências entre pontos nas imagens de profundidade. Após a geração da nuvem de pontos e a remoção de *outliers*, os

parâmetros de transformação do registro são estimados, inicialmente, pelo método de Horn, e, refinados pela minimização de erros de reprojeção usando o MMQ não-linear. Entretanto, devido à baixa qualidade das imagens de profundidade, ruídos remanescentes podem influenciar negativamente no estabelecimento correto das correspondências. Em Dryanovsky *et al.* (2012) essas imagens também são utilizadas no estabelecimento de correspondências podendo gerar essas inconsistências. Ambos os trabalhos não se atentam ao correto sincronismo entre dados RGB e de profundidade. Segundo Dryanovsky *et al.* (2012), nessa abordagem *coarse-to-fine* para o registro de nuvens de pontos 3D obtidas com dados RGB-D, os resultados são significativamente melhores que a execução dos métodos por si próprios.

Endres *et al.*, (2012) propõe um método para mapeamento 3D de ambientes internos em que feições são extraídas e correspondidas nas imagens RGB, em seguida, essas correspondências são associadas a nuvem de pontos 3D e os parâmetros do registro dos pares de nuvens são estimados usando RANSAC. Por fim, a consistência global é alcançada com o método G²O. No entanto, os autores utilizam apenas do RANSAC e do critério de distância euclidiana mínima para lidar com o problema do não-sincronismo dos dados RGB e profundidade.

Em Henry *et al.* (2012), basicamente, quatro pontos principais são discutidos: Primeiramente, pontos visuais são detectados e correspondidos por meio do algoritmo SIFT (*Scale Invariant Feature Transform*) idealizado por Lowe (1999), enquanto o estabelecimento de correspondências entre pontos nas duas imagens RGB é feito através de uma análise de distância entre os candidatos a correspondência; em seguida, é feita uma associação entre os pares de pontos correspondentes na imagem RGB e seus respectivos pontos contidos nos pares de nuvens de pontos 3D. Em seguida, é realizado o registro dos pares de nuvens de pontos 3D. Esta etapa é realizada através do algoritmo ICP; então, uma vez que cada par de nuvens de pontos 3D é registrado em um sistema referencial local e os erros obtidos no processo de registro são linearmente propagados, os autores empregaram um algoritmo de detecção de lugares revisitados (*loop closure*); e, por fim, é realizado o ajustamento global dos pares de nuvens de pontos independentes com o método TORO (*Tree-Based netwORk Optimizer*), proposto por Grisetti *et al.* (2009), refinando os parâmetros de registro obtidos e transformando o conjunto de dados em um único sistema referencial global.

Khoshelham *et al.* (2013) propuseram um método para o registro de pares de nuvens de pontos 3D obtidos com câmeras RGB-D baseado em abordagem ponto-a-ponto. Neste trabalho, os autores apresentaram um modelo de erros ponderados para melhorar a precisão da trajetória do sensor 3D. A ponderação é baseada no modelo teórico apenas do valor de profundidade. Eles também mostraram a necessidade de alinhar perfeitamente as imagens RGB e de profundidade para associar os pontos 2D e 3D correspondentes usando uma resseção espacial de imagens. O método proposto mostrou melhoras significativas na trajetória do sensor e a potencialidade do emprego de pesos nas observações. Porém, não é feita uma retificação da imagem IR em relação a imagem RGB no momento da associação dos pontos 2D para 3D. Além do mais, apenas a coordenada Z está sujeita à ponderação, sem aplicação ou nenhuma investigação mais aprofundada do uso de pesos em X e Y .

Weber *et al.* (2015) propõe um método que alinha automaticamente todas as nuvens de pontos em um sistema de coordenadas global realizando uma sequência de sucessivos registros dos pares de nuvens de pontos. Inicialmente, descritores de pontos são calculados na nuvem para o estabelecimento de correspondências por meio do FPFH (Fast Point Feature Histograms – RUSU *et al.*, 2009), em seguida, um grafo global com a nuvem de pontos é calculado, então, por heurística, são selecionados os pares de nuvens de pontos a serem alinhados. Para o registro dos pares de nuvens de pontos, são encontrados os parâmetros para um alinhamento inicial das nuvens por uma transformação de corpo rígido solucionada por SVD (Singular Value Decomposition – Arun *et al.*, 1987), em seguida, esses parâmetros são refinados com o ICP. Dada a propagação de erros advinda dos sucessivos registros de pares de nuvens de pontos 3D, os autores usaram para o ajustamento global uma adaptação do método de LUM proposta em Borrmann *et al.* (2008). Nessa variante, as posições das nuvens de pontos 3D são otimizadas com 6 graus de liberdade e não apenas com 3 como em Lu e Millios (1997), assim, uma linearização, por expansão de Taylor, da rotação deve ser feita. Entretanto, o método proposto por Weber *et al.* (2015) falha quando uma nuvem de pontos 3D não possua qualquer sobreposição com as demais nuvens de pontos.

Neste trabalho, diferente de Borrmann *et al.* (2008), os parâmetros de translação entre os pares de nuvens de pontos 3D são inseridos diretamente no ajustamento global, sem a necessidade de qualquer linearização.

2.2. DISPOSITIVO KINECT

Em novembro de 2010 a Microsoft lançou o dispositivo Kinect para fins de entretenimento e jogos computacionais da Microsoft Xbox 360. Este dispositivo foi desenvolvido pela empresa *PrimeSense* em colaboração com a Microsoft. O Kinect é um equipamento leve, flexível e de baixo custo quando comparado com sistemas de varredura LASER e câmeras de distância 3D. A Figura 1 apresenta um dispositivo Kinect e seus sensores integrados. Como pode ser visto na Figura 1, o dispositivo Kinect é composto, basicamente, por três sensores: dois sensores CMOS (*Complementary Metal-Oxide-Semiconductor*) que registram energia eletromagnética na faixa do espectro correspondente ao visível (RGB - *Red Green e Blue*) e infravermelho próximo (IR - *InfraRed*); e um emissor LASER infravermelho; além de microfones e um sensor de rotação motorizada. Os sensores RGB e IR capturam cenas com 640 x 480 pixels em uma taxa de 30 quadros por segundo (fps), sendo que cada quadro capturado contém até 300 mil pontos. Devido a esta característica, o Kinect pode propiciar uma nuvem de pontos 3D colorida com alta riqueza de informações visuais.

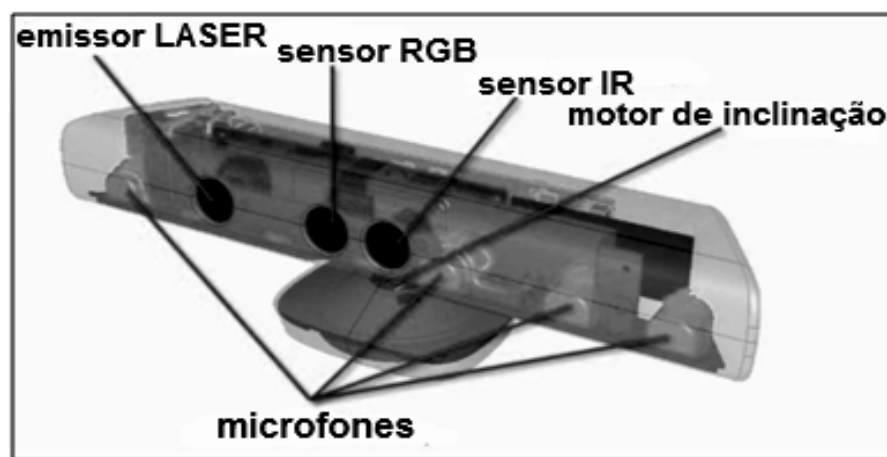


FIGURA 1 - SENSORES INTEGRANTES DO KINECT

FONTE: <http://kinectando.com.br>

De acordo com Freedman *et al.* (2010) o princípio de funcionamento do dispositivo Kinect para determinação dos valores de profundidade é conhecido como triangulação 3D. Basicamente, o sensor LASER emite um pulso na região do espectro correspondente ao infravermelho e um padrão de luz estruturada é criado a

partir de um processo de difração da luz. A Figura 2 mostra um exemplo de imagem do padrão de luz estruturada criado.



FIGURA 2 - PADRÃO DE LUZ ESTRUTURADA GERADO PELO KINECT
FONTE: O Autor (2015)

O padrão de luz estruturada é usado para calcular a distância entre o sensor e a superfície física de referência (Z_0 , ver Fig. 3) que é, previamente, armazenada na memória do dispositivo. Em outras palavras, é criado um plano de referência (linha pontilhada na Fig. 3). Um ponto (o) qualquer do padrão de luz estruturada e contido no plano de referência é registrado na imagem IR com sua posição deslocada (ver o ponto k na Fig. 3). Para resolver o problema de correspondência, cada pixel na imagem IR é comparado com os pixels do padrão de luz estruturada (armazenado anteriormente) através de uma convolução em torno da sua vizinhança, cujo tamanho da janela de correlação pode variar de 7×7 , 9×9 ou 9×7 . O pixel na imagem IR apresentando a maior correspondência dentro da janela de correlação é selecionado como o pixel correspondente. O deslocamento entre o pixel de referência do padrão de luz estruturada e seu correspondente na imagem IR indica o valor de paralaxe (d). Desta forma, o valor de profundidade de cada pixel na imagem IR pode ser obtido em função de d , da distância focal do sensor IR, da linha de base (b) e de Z_0 (ver Equação 3). A Figura 3 apresenta a geometria do modelo matemático, proposto por Khoshelham e Elberink (2012), para calcular os valores de profundidade Z_k e as coordenadas X_k , Y_k de cada pixel na imagem IR.

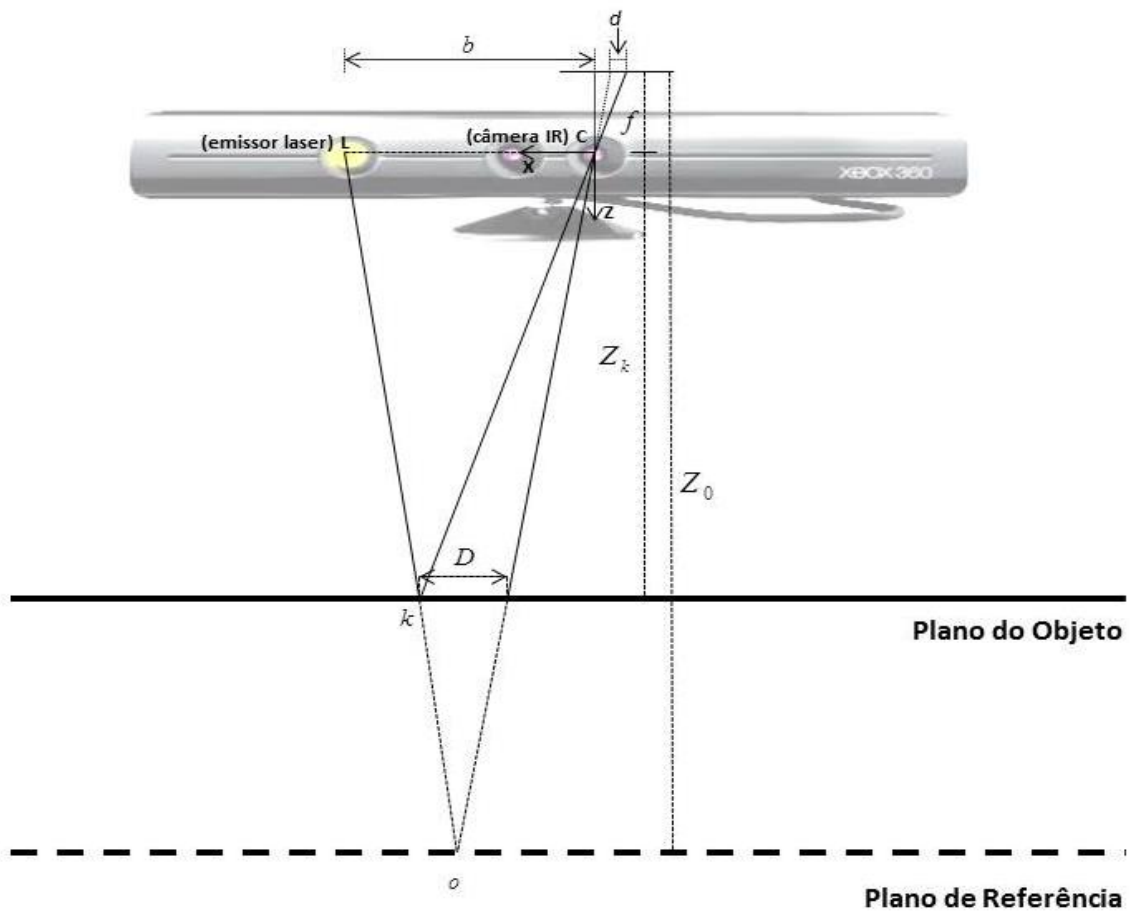


FIGURA 3 - ANALOGIA ENTRE PROFUNDIDADE RELATIVA E A PARALAXE MEDIDA

FONTE: O Autor (2015)

Assumindo que um objeto esteja posicionado entre o plano de referência e o sensor IR, um ponto capturado será transladado na direção X em relação ao ponto correspondente armazenado na memória do dispositivo. Como pode ser visto na Figura 3, este deslocamento corresponde à paralaxe observada no posicionamento do ponto k . Por semelhança de triângulos tem-se (KHOSHELHAM e ELBERINK, 2012):

$$\frac{D}{b} = \frac{Z_0 - Z_k}{Z_0} \quad (1)$$

$$\frac{d}{f} = \frac{D}{Z_k} \quad (2)$$

Substituindo a Equação (1) na (2) é possível expressar Z_k da seguinte forma:

$$Z_k = \frac{Z_0}{1 + \frac{Z_0}{f \cdot b} d} \quad (3)$$

A Equação (3) é o modelo matemático básico para a determinação da profundidade de um ponto pertencente à imagem observando sua paralaxe e assumindo que os parâmetros Z_0 , f e b são determinados por técnicas de calibração. Khoshelham e Elberink (2012) expressaram d em função de uma normalização linear. Neste caso, d (na Equação 3) é reescrito na forma $d = c_0 \cdot d' + c_1$, sendo c_0 e c_1 os parâmetros de uma normalização linear (na verdade uma denormalização) e a paralaxe normalizada. Substituindo os termos apresentados na Equação (3) e fazendo as devidas manipulações matemáticas, tem-se:

$$Z_k^{-1} = \left(\frac{c_0}{f \cdot b} \right) d' + \left(Z_0^{-1} \frac{c_1}{f \cdot b} \right) \quad (4)$$

A Equação (4) expressa uma relação linear entre o inverso da profundidade de um ponto e seu correspondente valor de paralaxe normalizado (KHOSHELHAM e ELBERINK, 2012). A coordenada Z_k em conjunto com f definem a escala da imagem para este ponto. As coordenadas planimétricas para cada ponto do objeto podem então ser calculadas a partir de suas coordenadas na imagem como segue (KHOSHELHAM e ELBERINK, 2012):

$$X_k = \frac{Z_k}{f} \cdot (x' - x_0 + \delta x) \quad (5)$$

$$Y_k = \frac{Z_k}{f} \cdot (y' - y_0 + \delta y) \quad (6)$$

onde, x' e y' são as coordenadas do ponto na imagem (observadas no plano focal da câmera IR), x_0 e y_0 são as coordenadas do ponto principal e δx , δy são as correções de distorção das lentes.

As Equações (4)-(6) também são conhecidas como equações de paralaxe e podem ser usadas para calcular os pontos no espaço-objeto (HALLERT, 1960; MOFFIT e MIKHAIL, 1980).

2.2.1. Calibração de Sensores

O processo de calibração consiste em determinar os POI dos sensores RGB e IR que compõem o dispositivo Kinect e os POE de ambas as câmeras. Os POI são determinados a partir da seguinte equação (ZHANG, 2000):

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} fS_x & \tau & x_0 \\ 0 & fS_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (7)$$

sendo, x' e y' as coordenadas do ponto no espaço-imagem (p'), X, Y, Z as coordenadas do ponto correspondente no espaço-objeto (P), x_0 e y_0 as coordenadas do ponto principal, τ o fator de não-ortogonalidade entre os eixos da câmera, f a distância focal da câmera, S_x e S_y os fatores de escala em x e y , $r_{i,j}$ corresponde aos elementos da matriz de rotação em função dos ângulos de Euler (ω, ϕ, κ) com $i=j=1, \dots, 3$, t_x, t_y, t_z são os parâmetros de translação e $' \cdot '$ é o operador de multiplicação.

Ainda é necessário considerar os coeficientes de distorção do sistema de lentes da câmera (distorção radial e descentrada), como segue (BROWN, 1971):

$$\begin{aligned} \delta x &= \bar{x} \cdot (k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) + P_1 \cdot (r^2 + 2 \cdot \bar{x}^2) + 2 \cdot P_2 \cdot \bar{x} \bar{y}^2 \\ \delta y &= \bar{y} \cdot (k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) + P_2 \cdot (r^2 + 2 \cdot \bar{y}^2) + 2 \cdot P_1 \cdot \bar{x} \bar{y}^2 \end{aligned} \quad (8)$$

sendo, $\bar{x} = x' - x_0$, $\bar{y} = y' - y_0$, $r = \sqrt{\bar{x}^2 + \bar{y}^2}$, k_1, \dots, k_3 os coeficientes de distorção radial das lentes e P_1, P_2 os termos da distorção descentrada.

A relação entre os pontos no espaço-imagem e seus correspondentes no espaço-objeto é obtida usando um plano padrão. Usualmente, um tabuleiro de xadrez é empregado como campo de calibração e, por este motivo, a coordenada Z é inserida como injunção com valor igual a 0 (zero). A injunção imposta define uma transformação projetiva entre o campo de calibração e o plano da imagem. Desta

forma, o ponto $p' (x'_c, y'_c, 1)$ no espaço-imagem e seu correspondente $P (X, Y, 1)$ no espaço-objeto são representados por coordenadas homogêneas, como segue (ZHANG, 2000):

$$s \cdot p' = H \cdot P \quad (9)$$

sendo, s o fator de escala, $H = A \cdot [r_1 \ r_2 \ t]$ a matriz homográfica (HARTLEY e ZISSERMAN, 2003), A o primeiro termo do lado direito da Equação (7), r_1, r_2 , os elementos da primeira e segunda coluna da matriz de rotação e t o vetor de translação. Por comodidade de notação, P que antes era $X, Y, Z, 1$, passa a ser $X, Y, 1$.

A partir da Equação (9), os POE são, simultaneamente, determinados segundo a matriz H . Fazendo $H = [h_1, h_2, h_3]$, obtém-se (ZHANG, 2000):

$$H = [h_1, h_2, h_3] = s \cdot A [r_1, r_2, t] \quad (10)$$

Como r_1 e r_2 são vetores ortogonais, obtém-se duas restrições básicas nos parâmetros intrínsecos, dada uma homografia (ZHANG, 2000):

$$h_1^T \cdot A^{-T} \cdot A^{-1} \cdot h_2 = 0 \quad (11)$$

$$h_1^T \cdot A^{-T} \cdot A^{-1} \cdot h_1 = h_2^T \cdot A^{-T} \cdot A^{-1} \cdot h_2 \quad (12)$$

Seja $B = A^T \cdot A^{-1}$ uma matriz simétrica, é possível definir os elementos de B pelo vetor $bz = [B_{11} \ B_{12} \ B_{22} \ B_{13} \ B_{23} \ B_{33}]^T$. Seja o i -ésimo vetor coluna de H $h_i = [h_{i1}, h_{i2}, h_{i3}]^T$, então chegamos a:

$$h_i^T \cdot B \cdot h_j = v_{ij}^T \cdot bz \quad (13)$$

onde $v_{ij} = [h_{i1}h_{j1}, h_{i1}h_{j2} + h_{i2}h_{j1}, h_{i2}h_{j2}, h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3}]^T$.

Assim, a partir das Equações (11) e (12), para cada imagem capturada obtêm-se duas equações homogêneas, como segue (ZHANG, 2000):

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{bmatrix} \cdot bz = 0 \quad (14)$$

De acordo com Zhang (2000), para n imagens capturadas, resolver a Equação (14) significa solucionar um sistema de equações do tipo $V \cdot bz = 0$, sendo V uma matriz de dimensões $2n \times 6$. A solução para a Equação (14) pode ser encontrada pelo método SVD. Encontrado os valores do vetor bz , é possível determinar os POI contidos na matriz A e o fator de escala s segundo as equações ilustradas no Quadro (1):

v_0	$\frac{(B_{12} \cdot B_{13} - B_{11} \cdot B_{23})}{(B_{11} \cdot B_{22} - B_{12}^2)}$
u_0	$\frac{\tau \cdot v_0}{f \cdot S_y} - \frac{B_{13} \cdot f \cdot S_x^2}{\tau}$
s	$\frac{B_{33} - (B_{13}^2 + v_0(B_{12} \cdot B_{13} - B_{11} \cdot B_{23}))}{B_{11}}$
fS_x	$\sqrt{\frac{s}{B_{11}}}$
fS_y	$\sqrt{\frac{s \cdot B_{11}}{B_{11} \cdot B_{12} - B_{12}^2}}$
τ	$-B_{12} \cdot f \cdot S_x^2 \frac{f \cdot S_y}{s}$

QUADRO 1 - DETERMINAÇÃO DOS PARÂMETROS DE ORIENTAÇÃO INTERIOR
FONTE: ZHANG (2000)

Os POE da câmera são obtidos fazendo (ZHANG, 2000):

$$r_1 = \lambda \cdot A^{-1} \cdot h_1 \quad (15)$$

$$r_2 = \lambda \cdot A^{-1} \cdot h_2 \quad (16)$$

$$r_3 = r_1 \otimes r_2 \quad (17)$$

$$t = \lambda \cdot A^{-1} \cdot h_3 \quad (18)$$

sendo, $\lambda = \frac{1}{\|A^{-1} \cdot h_1\|} = \frac{1}{\|A^{-1} \cdot h_2\|}$ e \otimes o operador de produto vetorial. Vale ressaltar que a melhor solução para a matriz de rotação pode ser encontrada em Golub e Van Loan (1996).

2.3. REGISTRO DOS PARES DE NUVENS DE PONTOS 3D

Basicamente, o processo de registro consiste em determinar os parâmetros de transformação entre pares de nuvens de pontos 3D visando estabelecê-los em um único sistema referencial. Isto garante condições de mapeamento entre qualquer ponto no sistema referencial de coordenadas da nuvem de referência e seus pontos correspondentes na nuvem de pesquisa e vice-versa (SIMON, 1996). Chen e Medioni (1991) afirmaram que duas nuvens de pontos são ditas registradas quando uma delas é posicionada e orientada corretamente em relação à outra. Basicamente, esta solução somente pode ser encontrada se for estabelecido um número suficiente de pontos correspondentes e através de um modelo de otimização e um método de estimativa de parâmetros, tal como o MMQ.

No contexto atual de mapeamento 3D de ambientes internos, os métodos mais robustos de registro de pares de nuvens de pontos dividem a tarefa em duas etapas: 1) Determinação dos parâmetros de transformação do registro grosseiro; 2) Refinamento desses parâmetros.

2.3.1. Determinação dos parâmetros de transformação do registro grosseiro

Com o surgimento de câmeras RGB-D há uma sinergia natural entre pontos visuais (detectados na imagem RGB) e valores de profundidade que pode ser explorada no contexto do registro de pares de nuvens de pontos 3D. Para isto, é necessário detectar e extrair primitivas pontuais no par de imagens RGB. O método mais conhecido na comunidade fotogramétrica é o algoritmo SIFT.

Para a execução do algoritmo SIFT é necessário o reconhecimento de objetos ou detecção de feições locais de imagem independente da complexidade e

desordem natural presente nos ambientes internos. Dessa forma, essas feições detectadas precisam ser suficientemente distintas para identificar objetos específicos entre muitas alternativas, além do mais, devem ser pelo menos parcialmente invariantes quanto à iluminação, ruído de imagem, transformações projetivas 3D e as variações de objetos comuns (LOWE, 1999). Para identificar feições locais em imagens o algoritmo SIFT utiliza descritores locais invariantes, que possibilitam ao algoritmo encontra-las de uma forma recorrente (os mesmos descritores extraídos de duas imagens devem representar o mesmo objeto) e as codificar em uma representação que seja invariante a quaisquer transformações.

Extraídas as primitivas locais, os descritores podem ser definidos como conjuntos de medidas que representam a essência dessas primitivas, ou seja, é uma “assinatura” do objeto detectado.

No método apresentado por Lowe (1999), uma imagem é transformada em uma grande coleção de descritores locais, cada um dos quais é invariante a translação, a escala, a rotação, e parcialmente invariante a mudanças de iluminação e de projeção afim ou 3D. A principal ideia deste algoritmo é reduzir o custo computacional com tempo de processamento através de operações executadas apenas em pontos potenciais (regiões com forte variação de níveis de cinza, tais como bordas ou cantos). Basicamente, o algoritmo SIFT possui três etapas bem definidas: detecção de pontos de interesse; a geração de descritores; e o estabelecimento de correspondências.

Primeiramente, o algoritmo procura pontos invariantes à mudança de escala na imagem utilizando uma função Gaussiana (espaço-escala). O espaço-escala é definido como uma convolução da imagem $I(u, v)$ com um operador Gaussiano $G(u, v, \vartheta)$, como segue (LOWE, 2004):

$$L(u, v, \vartheta) = G(u, v, \vartheta) \otimes I(u, v) \quad (19)$$

sendo, L a imagem suavizada, u e v as coordenadas do pixel na imagem RGB, ϑ a variação de escala em G (parâmetro que define o fator de suavização da imagem I)

$$\text{e } G(u, v, \vartheta) = \frac{1}{2\pi\vartheta^2} e^{-\frac{(u^2+v^2)}{2\vartheta^2}}.$$

Em seguida, são realizados cálculos de Diferença de Gaussianas (*Difference of Gaussians* - DoG) para cada oitava (quando o valor da variação ϑ aumenta) no

espaço-escala. Esta função é separada por uma escala constante (k) e definida pela diferença de L em escalas próximas (ver Figura 4), como segue (LOWE, 2004):

$$DoG(u, v, \mathcal{S}) = L(u, v, k \cdot \mathcal{S}) - L(u, v, \mathcal{S}) \quad (20)$$

A Figura 4 ilustra duas oitavas, estando à esquerda o conjunto de imagens suavizadas pelos filtros Gaussianos e à direita o conjunto de imagens resultantes da Diferença de Gaussianas.

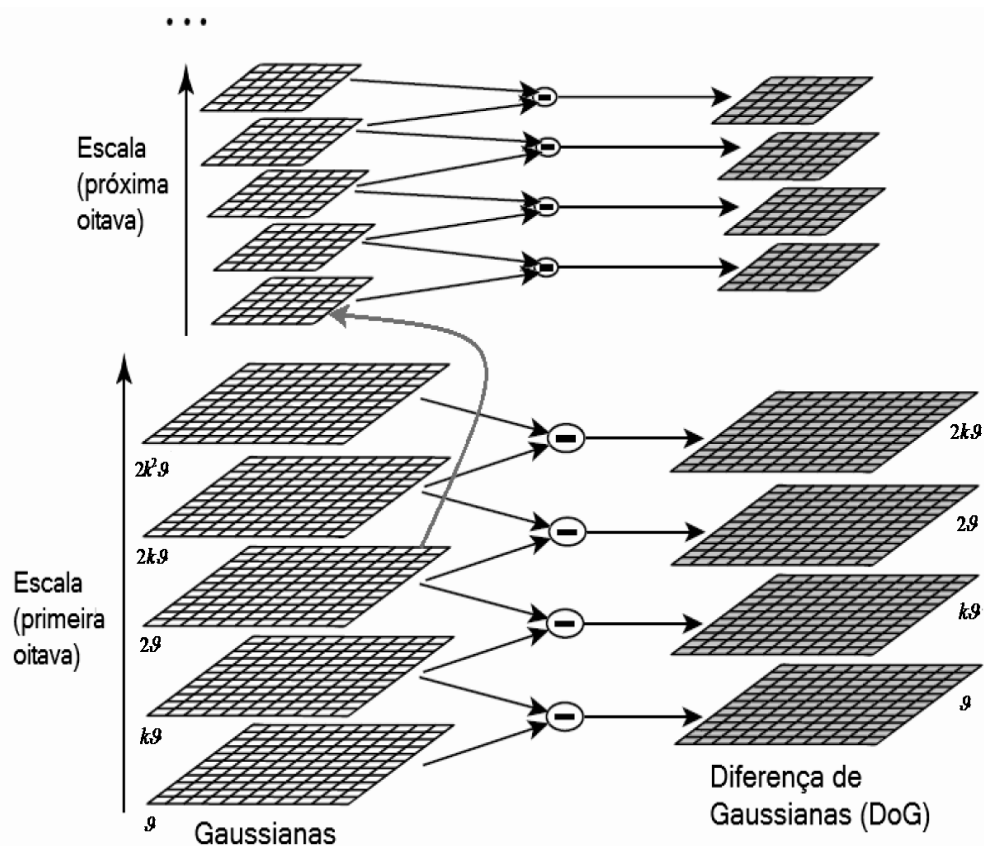


FIGURA 4 - REPRESENTAÇÃO DO PROCESSO PARA CONSTRUÇÃO DAS IMAGENS RESULTANTES DA DIFERENÇA DE GAUSSIANS EM CADA OITAVA NO ESPAÇO ESCALA

FONTE: ADAPTADO DE LOWE (2004)

Esse conceito de utilizar uma função gaussiana no espaço escala para este propósito é baseado no trabalho de Lindeberg (1994). A Equação (20) proporciona um valor aproximado para a escala normalizada Laplaciana-Gaussiana ($\mathcal{S}^2 \nabla^2 G$), como segue (LOWE, 2004):

$$\vartheta^2 \nabla^2 G = \frac{\partial G}{\partial \vartheta} \approx \frac{L(u, v, k \cdot \vartheta) - L(u, v, \vartheta)}{k \cdot \vartheta - \vartheta} \quad (21)$$

e, por conseguinte:

$$L(u, v, k \cdot \vartheta) - L(u, v, \vartheta) \approx (k - 1) \cdot \vartheta^2 \nabla^2 G \quad (22)$$

O fator ϑ^2 faz com que a escala seja invariante. Na sequência, a localização dos pontos visuais é feita por aproximações de $\vartheta^2 \nabla^2 G$. Conforme mostrado por Lowe (2004), o fator remanescente $(k - 1)$ não influencia na posição do pixel detectado. O emprego da função gaussiana tem o objetivo de obter imagens com ruídos atenuados e bordas realçadas. O resultado esperado é um conjunto de quatro oitavas, sendo realizadas cinco operações gaussianas para cada oitava.

Na sequência, os valores de máximo ou mínimo locais (pontos visuais extremos) devem ser detectados para cada *DoG* obtido. Este procedimento é baseado na comparação de todos os pixels de borda na imagem *DoG* com seus vizinhos (vizinhança 8), do espaço-escala corrente, em relação aos vizinhos correspondentes no espaço-escala adjacentes superior e inferior (ver Figura 5). Na Figura 5, o ponto marcado com “X” é comparado com seus vizinhos marcados como “O”.

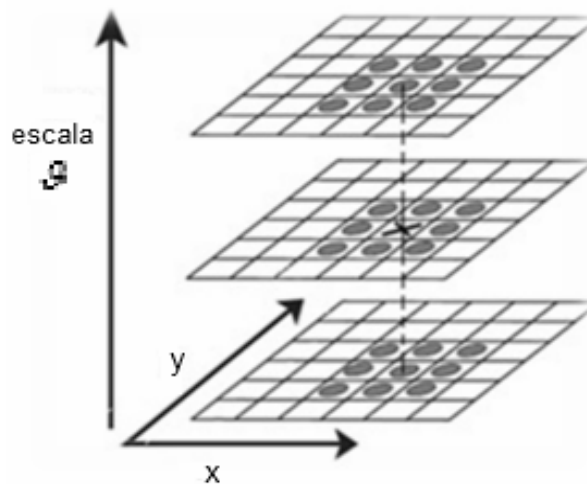


FIGURA 5 - EXEMPLO DO PROCESSO DE DETECÇÃO DE EXTREMOS NO ESPAÇO-ESCALA

FONTE: ADAPTADO DE LOWE (2004)

Posteriormente, os pontos visuais devem ser localizados e se não forem estáveis, devem ser descartados. A localização exata dos pontos visuais é determinada através do ajustamento de uma função quadrática 3D (BROW E LOWE, 2002). Dado um ponto de amostragem, sua localização interpolada do máximo é determinada através da expansão da série de Taylor da função *DoG*, a saber (LOWE, 2004):

$$DoG(\Delta) = DoG + \frac{\partial DoG}{\partial \Delta} \Delta + \frac{1}{2} \Delta^T \cdot \frac{\partial^2 DoG}{\partial \Delta^2} \Delta \dots \quad (23)$$

onde $\Delta = (u, v, \vartheta)^T$.

Na Equação (23) a imagem $DoG(u, v, \vartheta)$ é deslocada (Δ) de tal forma que sua origem esteja localizada no ponto de amostragem. Enfim, a localização do extremo (ponto visual) pode ser calculada derivando a Equação (23) em função de (Δ) e igualando o resultado à zero, como segue (LOWE, 2004):

$$\Delta = - \frac{\partial^2 DoG}{\partial \Delta^2}^{-1} \cdot \frac{\partial DoG}{\partial \Delta} \quad (24)$$

Na sequência, para cada ponto visual é atribuída uma orientação canônica fazendo com que os descritores sejam invariantes à rotação (LOWE, 1999). Para cada imagem $L(u, v, \vartheta)$, na mesma escala ϑ , a magnitude do gradiente mag_{uv} e a orientação α_{uv} de cada ponto visual são computadas usando diferenças de pixels, conforme as equações a seguir (GONZALES e WOODS, 2000).

$$mag_{uv} = \sqrt{(L(u+1, v) - L(u-1, v))^2 + (L(u, v+1) - L(u, v-1))^2} \quad (25)$$

$$\alpha_{uv} = \tan^{-1} \left(\frac{L(u, v+1) - L(u, v-1)}{L(u+1, v) - L(u-1, v)} \right) \quad (26)$$

Para cada ponto visual e seus pixels vizinhos é gerado um histograma de orientações. Os pixels vizinhos são adicionados ao histograma com pesos correspondentes aos valores de mag_{uv} e também aos valores determinados por uma função Gaussiana. Esta função Gaussiana é determinada por uma janela de suavização Gaussiana com σ igual à metade da largura da janela do descritor da

função. O valor dos pesos atualizados para cada pixel vizinho é dado por (LOWE, 2004):

$$\omega g' = \omega g + \alpha_{uv} \text{mag}_{uv} \cdot Gs(\Delta u, \Delta v, \vartheta') \quad (27)$$

Com:

$$\alpha_{uv} = \begin{cases} dtn/i, & dtn < i \\ 0, & dtn > i \end{cases}$$

sendo, $\omega g'$ a atualização de ωg , $Gs(\Delta u, \Delta v, \vartheta')$ a janela Gaussiana, dtn a distância absoluta em graus entre a orientação do ponto e o θ discretizado, e i o intervalo em graus entre os $\theta's$ discretizados.

O histograma possui 36 regiões, cobrindo todas as orientações possíveis (intervalos de 10°). A Figura 6 mostra um exemplo de um ponto e seus pixels vizinhos com suas orientações e magnitudes, e, o histograma dessas orientações.

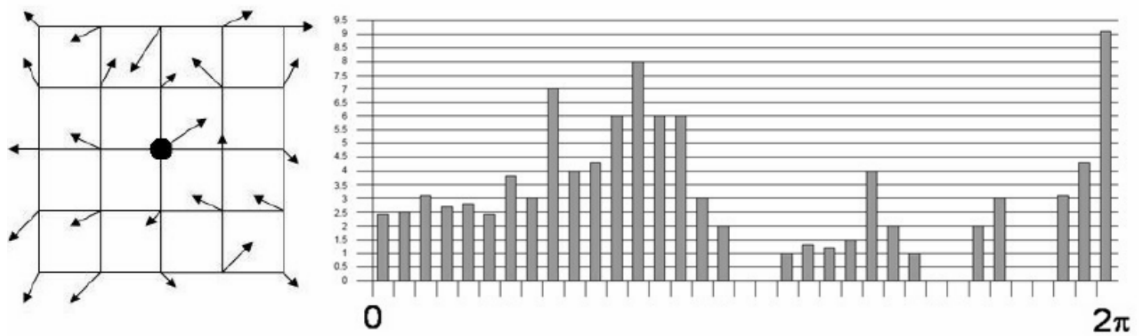


FIGURA 6 - HISTOGRAMA DE ORIENTAÇÕES DE UM PONTO VISUAL

FONTE: www.maxwell.vrac.puc-rio.br

Picos na orientação do histograma correspondem a direções dominantes para os gradientes locais. O maior pico detectado no histograma e os picos que correspondem a valores percentuais acima de 80% do maior pico são usados para definir a orientação de cada ponto visual. Então, os valores obtidos são normalizados e para cada ponto visual é atribuído um descritor invariante à iluminação, ao ponto de vista e à rotação. O descritor consiste de um vetor contendo os valores do histograma com n elementos determinados em função de uma janela de apoio com um total de $j \times j$ pixels (ver Figura 7).

A Figura 7 mostra uma matriz ($j = 2$) de orientação de histogramas com 8 direções em cada posição, representando um vetor ($nu = 32$) com elementos descritores para cada ponto visual.

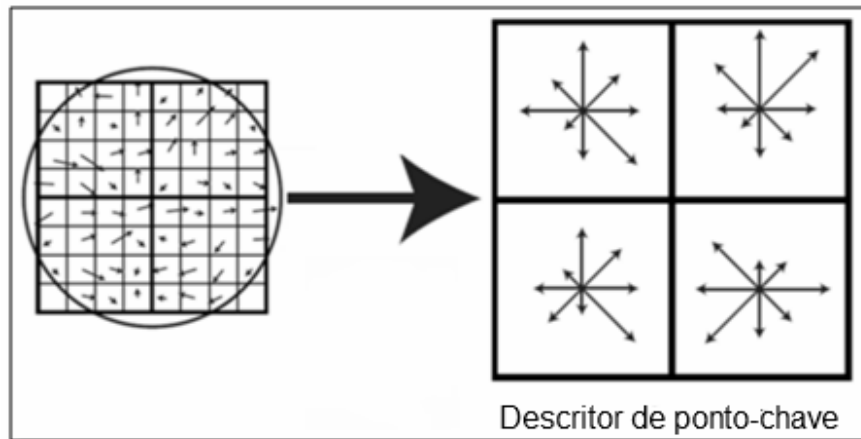


FIGURA 7 - CONSTRUÇÃO DO DESCRITOR PARA UM PONTO VISUAL

FONTE: ADAPTADO DE LOWE (2004)

O resultado esperado é um conjunto de descritores que podem ser usados no estabelecimento automático de correspondências entre pontos. Finalmente, a correspondência entre os pontos de interesse pode ser efetuada por intermédio do cálculo de alguma medida de similaridade que descreve a relação entre os descritores em uma dada vizinhança.

Segundo Lowe (2004), o melhor candidato a correspondência para cada ponto visual é encontrado pela identificação de seu vizinho mais próximo na outra imagem. Beis e Lowe (1997) utilizaram uma modificação do algoritmo *k-d tree* (*k-dimensional tree*) chamado de método de *Best-Bin-First* (BBF), identificando os vizinhos mais próximos de alta probabilidade, utilizando apenas uma quantidade limitada de esforço computacional.

2.3.2. Geração da Matriz Fundamental por RANSAC

Embora o algoritmo SIFT detecte uma grande quantidade de pontos visuais, existem diversos falsos positivos, tratados como *outliers*, que devem ser detectados e removidos do processo. Segundo Bradski e Kaebler (2008) é possível através de

uma geometria epipolar detectar e eliminar falsas correspondências, desde que a distância perpendicular entre a reta epipolar originada de um ponto visual na imagem da esquerda e seu correspondente na imagem da direita seja maior que um determinado limiar. Assim, a correspondência estabelecida por esses pontos será removida uma vez que será considerada como falsa correspondência.

De acordo com Feng e Hung (2003), a geometria epipolar pode ser expressa em termos da Matriz Fundamental (MF) que contém todas as informações geométricas (POI e POE) necessárias para o estabelecimento da correspondência entre pares de pontos detectados em imagens sucessivas. Basicamente, a MF relaciona os pares de pontos correspondentes x_i e x'_i , sendo x_i um ponto visual qualquer pertencente à imagem de referência e x'_i o possível ponto correspondente na imagem de pesquisa (ambas corrigidas dos erros sistemáticos), de forma que a equação a seguir seja satisfeita (ZHANG, 1998):

$$x_i'^T \cdot F \cdot x_i = 0 \quad (28)$$

Sendo F a Matriz Fundamental de ordem 3 com seu determinante nulo e que pode ser calculada da seguinte forma:

$$x_i'^T \cdot F \cdot x_i = \begin{bmatrix} x'_i & y'_i & 1 \end{bmatrix} \begin{bmatrix} cf_{11} & cf_{12} & cf_{13} \\ cf_{21} & cf_{22} & cf_{23} \\ cf_{31} & cf_{32} & cf_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0 \quad (29)$$

sendo, cf_{11}, \dots, cf_{33} os coeficientes de F .

Fazendo as devidas multiplicações matemáticas e considerando, $cf = (cf_{11}, cf_{12}, cf_{13}, cf_{21}, cf_{22}, cf_{23}, cf_{31}, cf_{32}, cf_{33})$, a Equação (29) pode ser reescrita da forma, a saber:

$$cf(x'_i x_i, x'_i y_i, x'_i, y'_i x_i, y'_i y_i, y'_i, x_i, y_i, 1) = 0 \quad (30)$$

Assim, para solucionar a Equação (30), podem ser usadas técnicas como SVD, MMQ e etc. Para testar quais as melhores correspondências na geração da

Matriz Fundamental, pode ser empregado o RANSAC (FISCHLER e BOLLES, 1981). O RANSAC é baseado na noção de conjuntos mínimos, que consiste no menor número de amostras necessárias para, de maneira unívoca, definir um modelo. Basicamente, o algoritmo se divide em: 1), estabelecer uma hipótese (um conjunto mínimo de correspondências é aleatoriamente selecionado a partir de todo conjunto de dados disponível); 2) num teste (o algoritmo verifica quais elementos de todo conjunto de dados que são consistentes com o modelo estabelecido pelo conjunto mínimo). Os elementos considerados consistentes formarão um conjunto chamado de consenso. Esses dois processos ocorrem iterativamente, e, ao final deles, o conjunto que será adotado como modelo corresponderá ao conjunto mínimo que apresentar o maior conjunto consenso.

Assim, para um conjunto de correspondências com uma respectiva porcentual γ de *inliers*, o número de tentativas nt necessárias para suficientemente oferecer uma probabilidade máxima Pz na escolha de um conjunto sem *outliers* consistindo de nc correspondências entre pontos é dada por (FENG e HUNG, 2003):

$$nt = \frac{\log(1-Pz)}{\log(1-(1-\gamma)^{nc})} \quad (31)$$

Então, todo o processo de construção da MF por meio do RANSAC pode ser dividido nos seguintes passos (FENG e HUNG, 2003):

1. Iteração nt vezes, onde nt é ajustado usando a equação acima quando γ é atualizado.
 - a) Seleção de um subconjunto mínimo S a partir do conjunto total de dados.
 - b) Cálculo de F_k usando S (MMQ, SVD etc)
 - c) Avaliação da consistência de F_k com todas as correspondências do conjunto total de dados.
 - i. Cálculo do erro quadrático $\{e_i\}$ para cada ponto do conjunto total de dados com F_k ;
 - ii. Avaliação de F_k por uma função de *score* separando o conjunto de dados em *inliers* e *outliers*.

2. Escolha da melhor solução ao longo de um conjunto de soluções $\{F_k\}$, $k = 1, \dots, N_s$ de acordo com seus *scores*.

A função *score* é empregada para avaliar se a Matriz Fundamental se encaixa com o conjunto de dados testado (*inliers*). Ao particionar o conjunto de dados em *inliers* e *outliers*, o algoritmo permite a eliminação dos valores que não estão adequados ao modelo. A função *score* do RANSAC é dada pela equação a seguir:

$$f(k) = \binom{t}{k} \cdot ps^k (1 - ps)^{nt-k} \quad (32)$$

onde k é o número de tentativas que resultam em sucesso e ps a probabilidade de sucesso em cada tentativa. Cabe dizer que a solução pode ser refinada através de um ajustamento pelo MMQ (ver seção 2.3.3) empregando os dados contidos no conjunto consenso.

Após os pontos visuais serem detectados e extraídos nos pares de imagens RGB, suas correspondências estabelecidas e removidos os *outliers*, é possível associá-los com seus pontos correspondentes na nuvem de pontos 3D atribuindo as profundidades correlatas. Assim, os parâmetros de transformação entre os pares de nuvens de pontos podem ser estimados através de uma transformação de corpo rígido 3D e pelo MMQ.

2.3.3. Modelo Paramétrico de Ajustamento

O ajustamento pelo MMQ é o ramo da matemática aplicada que tem por finalidade encontrar uma solução única para problemas onde o número de observação é maior que o seu numero de parâmetros (DALMOLIN, 2004). Nesse sentido, o MMQ consiste em estimar variáveis estocásticas X e seus parâmetros de distribuição Σ_x , a partir de amostras L observadas com precisão Σ_L .

De acordo com Dalmolin (2004), o modelo matemático de ajustamento paramétrico, também chamado de modelo explícito ou método das observações

indiretas, expressa as observações ajustadas em função dos parâmetros ajustados. Os modelos funcionais aplicados a esse método são formados por equações implícitas do tipo:

$$L_a = F(X_a) \quad (33)$$

sendo, L_a o vetor das observações ajustadas, X_a o vetor dos parâmetros ajustados e F um funcional que relaciona L_a a X_a .

Seja $L_a = L_b + V$, em que L_b é o vetor das observações e V é o vetor de resíduos, substituindo o primeiro membro e linearizando o segundo obtém-se:

$$L_b + V = F(X_0 + X) = F(X_0) + \left. \frac{\partial F}{\partial X_a} \right|_{X_a=X_0} * X \quad (34)$$

Designando a função dos parâmetros aproximados $L_0 = F(X_0)$ e a matriz das derivadas parciais $A = \left. \frac{\partial F}{\partial X_a} \right|_{X_a=X_0}$ tem-se:

$$L_b + V = L_0 + A \cdot X \rightarrow V = A \cdot X + L_0 - L_b \quad (35)$$

Fazendo $L = L_0 - L_b$ obtém-se o modelo linearizado do método paramétrico:

$$V = A \cdot X + L \quad (36)$$

Assim, para encontrar o vetor das correções dos parâmetros X faz-se:

$$X = (A^T \cdot P \cdot A)^{-1} A^T \cdot P \cdot L \quad (37)$$

sendo, P a matriz de pesos, então, o vetor dos parâmetros ajustados pode ser encontrado a partir da equação: $X_a = X_0 + X$. Na forma de ajustamento não-linear,

X_0 precisa de um valor inicial previamente determinado, e, o valor encontrado para X_a o substituirá na próxima iteração. Além do mais, pode ser usado como critério de parada do ajustamento tanto o valor absoluto de X (interrompe o processo quando atingir um valor menor que certo limiar) quanto se alcançar um determinado número de iterações.

2.3.4. Algoritmo ICP

No ICP, os parâmetros de transformação são obtidos através de uma função erro que minimiza o somatório da distância entre os pontos na nuvem de referência (M) e de pesquisa ($Pesq$). Apesar da eficiência na determinação dos parâmetros de transformação usando pares de nuvens de pontos 3D oriundas de qualquer sistema de imageamento, o algoritmo é computacionalmente lento, depende da entrada de valores iniciais aproximados e alta sobreposição entre os pares de nuvem de pontos. Basicamente, dado um par de nuvem de pontos 3D (nuvem de referência e nuvem de pesquisa) o algoritmo funciona da seguinte forma (Besl e Mckay, 1992):

1. São definidos os valores de rotação e translação iniciais. No caso desses valores não serem definidos, os valores adotados devem ser nulos;
2. Para cada ponto da nuvem de pesquisa são localizados os vizinhos mais próximos na nuvem de referência. Nessa etapa são estabelecidos os pares de pontos pseudo-conjugados. Para esse cálculo que localiza o vizinho mais próximo, deve ser considerada a distância métrica $dist$ entre um ponto individual q e o modelo M denotada por (BESL e MCKAY, 1992):

$$dist(Pesq, M) = \min_{m \in M} \|m - Pesq\| \quad (38)$$

sendo, $Pesq$ representado pelos Np pontos $\{p_i\}$ e M pelos Nm pontos $\{m_i\}$. Assim, o ponto mais próximo em M que produz a menor distância é denotado por x de modo que $dist(Pesq, x) = dist(Pesq, M)$, onde $x \in M$ (ver Figura 8).

Essa busca pode ser otimizada com o emprego da estrutura de dados conhecida como *K-d tree*. O método *K-d tree* (FRIEDMAN *et al.*, 1977) divide os

pontos na nuvem dependendo da sua localização, economizando cálculos de distâncias entre os candidatos a correspondências (SIMON, 1996). Basicamente, é calculada a média das coordenadas de todos os pontos da nuvem de pesquisa e se escolhe o ponto que mais se aproxima dessa medida. Esse ponto é considerado a “raiz” da *k-d tree* e é o ponto de partida para que sejam organizados todos os pontos dentro da árvore mantendo a relação de vizinhança entre eles.

3. Os parâmetros de transformação entre as duas nuvens de pontos são calculados pelo MMQ ou pelo método proposto por Horn (1987) e aplicados na nuvem de pesquisa;

4. Por fim, é calculada a Raiz do Erro Médio Quadrático (REMQ) entre as duas nuvens e comparado com o critério de convergência. Se o teste falhar, o processo é repetido novamente a partir do passo 2 entre a nuvem de pesquisa e de referência após a aplicação dos parâmetros calculados no passo 3. Caso contrário, o algoritmo encerra sua execução.

A Figura 8 mostra o processo de estabelecimento de pseudo-conjugados entre duas nuvens de pontos que representam a mesma superfície pelo método de vizinho mais próximo.

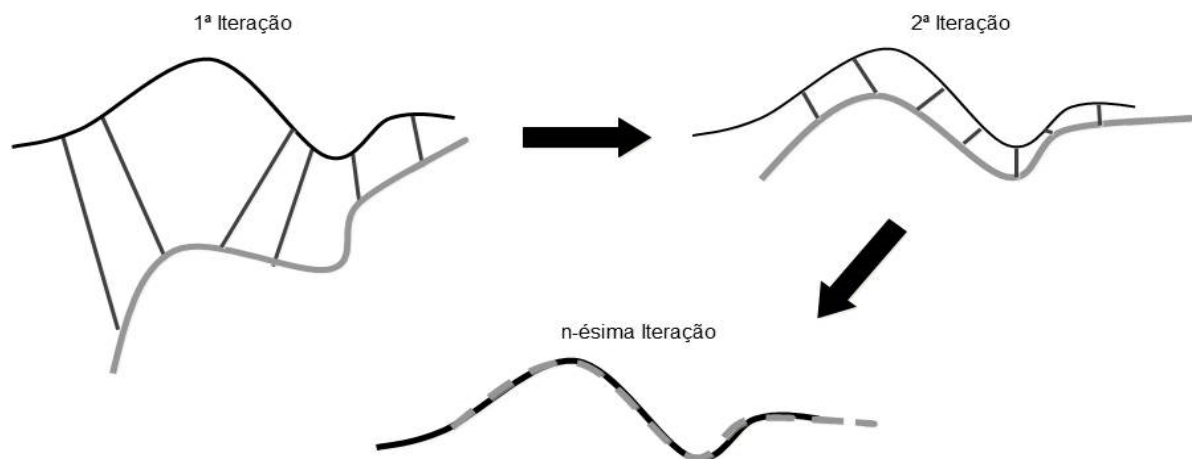


FIGURA 8 – PROCESSO DE ESTABELECIMENTO DE PSEUDO-CONJUGADOS ENTRE DUAS NUVENS QUE REPRESENTAM A MESMA SUPERFÍCIE PELO MÉTODO DE VIZINHO MAIS PRÓXIMO

FONTE: O autor (2015)

As principais vantagens dos métodos baseados no ICP consistem na simplicidade e no rápido desempenho do algoritmo quando implementados com a

busca de vizinho mais próximo por *k-d tree* e, principalmente, quando há um alinhamento inicial e alta sobreposição entre os pares de nuvens de pontos.

2.4. AJUSTAMENTO SEQUENCIAL DE PARES DE NUVENS DE PONTOS

De acordo com Williams e Brazier (1965), o ajustamento por modelos independentes é um problema clássico em Fotogrametria. Considerando que cada par de nuvens de pontos 3D é um modelo tridimensional da superfície física, neste trabalho, o problema será tratado como ajustamento sequencial de pares de nuvens de pontos.

Uma vez que os pares de nuvens de pontos 3D são independentes entre si (cada qual referido a um sistema referencial local) e os erros de alinhamento, de ruído, de quantização dos valores de profundidade, entre outros fatores, fazem com que a estimativa da posição da câmera derive ao longo do tempo levando a imprecisões em um SMAI-3D, seus parâmetros de transformação devem ser simultaneamente determinados. Basicamente, este método consiste em refinar, apenas, os parâmetros de translação de cada par de nuvens de pontos simultaneamente e transformar o mapa 3D em um sistema referencial global.

Para demonstrar o problema de propagação de erros ao longo da trajetória do sensor, será tomada como exemplo a Figura 9. Nesta figura, considere uma trajetória percorrida com um SMAI-3D começando no ponto K_0 , que corresponde ao primeiro *frame* adquirido naquela posição, e retornando para uma localização próxima K_n depois de visitar K_1, K_2, \dots, K_{n-1} , o novo conjunto de parâmetros é estimado simultaneamente baseado na seguinte equação proposta por Lu e Milios (1997), a saber:

$$F : X_{novo}^i = X_{j+1}^{i+1} - X_j^i \quad (39)$$

sendo, F o modelo funcional, X_{novo}^i o novo conjunto de parâmetros de translação ajustado, X_{j+1}^{i+1} a matriz homogênea dos parâmetros de translação da estação

posterior e X_j^i a matriz homogênea dos parâmetros de translação da estação anterior.

Neste caso, as observações são os valores de translação determinados na etapa de registro dos pares de nuvens de pontos 3D. Na Equação (39) cada estação ocupada pelo sensor está representada pelos parâmetros determinados em relação à estação atual e anterior, ou seja, $K_{ij} = [tx_{ij}, ty_{ij}, tz_{ij}]^T$, sendo $i = 0, \dots, K$ e $j = i+1$.

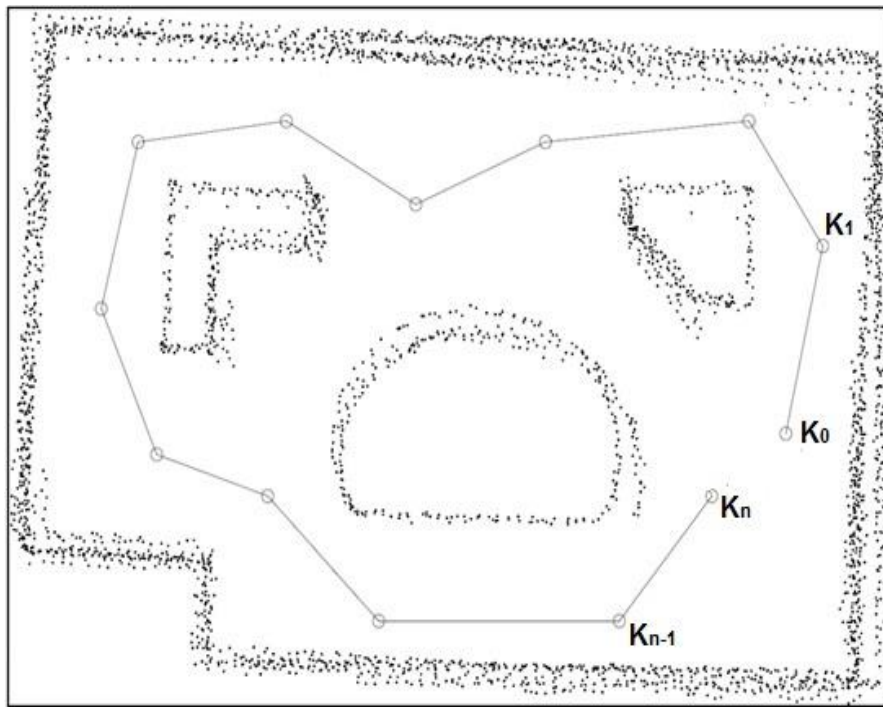


FIGURA 9 - POSES DESALINHADAS DEVIDO AO ACÚMULO DE ERROS
 FONTE: ADAPTADO DE LU e MILIOS (1997)

A seguir serão apresentados os materiais e métodos empregados neste trabalho.

3. MATERIAIS E MÉTODOS

No capítulo anterior foram apresentadas as características dos sensores embarcados no Kinect, bem como algumas técnicas para o registro dos pares de nuvens de pontos voltado ao mapeamento 3D de ambientes internos, além de uma ideia geral sobre o ajustamento por pares de nuvens de pontos independentes que pode ser realizado após o processo de registro de um conjunto de pares de nuvens de pontos 3D. Neste capítulo serão abordados os materiais e os métodos utilizados para a realização de todos esses procedimentos no âmbito da realização desta pesquisa.

3.3. MATERIAIS

3.3.1. Laboratório e Recursos Bibliográficos

- Laboratório de Aquisição e Processamento de Dados Espaciais (LAPE);
- Material bibliográfico disponível na biblioteca de Tecnologia e Ciências Exatas da Universidade Federal do Paraná;

3.3.2. Recursos de Hardware

- 01 Sensor RGB-D Kinect adquirido com recursos de bolsa de bancada processo no 400400/2013-8 linha 2 – Bolsa Pesquisador Visitante Especial PVE;
- 01 Laptop DELL *INSPIRION* N4050; Processador Intel Corei5; Sistema Operacional Windows 7 devidamente licenciado;

- 01 interferômetro da marca Hewlett Packard modelo 5508a pertencente ao Departamento de Geomática-UFPR.

3.3.3. Recursos de Software e Bibliotecas Computacionais *Open-Source*

- MRPT (*Mobile Robot Programming Toolkit*): Fornece aos desenvolvedores aplicativos e bibliotecas cobrindo estruturas de dados e algoritmos utilizados principalmente na robótica. É *open source*, distribuído sob a licença BSD (*Berkeley Software Distribution*). Um desses aplicativos é o kinect-estéreo-calib, que é um programa para calibração interativa dos parâmetros intrínsecos e extrínsecos dos sensores do Kinect.
- OpenCV (*Open-Source Computer Vision*): Biblioteca voltada a projetos de Visão Computacional e Processamento de Imagens; Linguagem C/C++, multiplataforma e *open-source*, distribuída sob licença BSD.
- Freemat: É um ambiente livre e de código aberto de análise numérica e, também, uma linguagem de programação similar ao MATLAB. Possui capacidades de plotagem e de visualização 3D.

3.4. MÉTODO

A Figura 10 mostra um diagrama do método proposto neste trabalho. Basicamente, o método está dividido em 4 etapas. Primeiramente, é realizada a calibração dos sensores embarcados no Kinect (RGB e IR) e determinados os parâmetros de montagem dos sensores. Em seguida, feito o registro de cada par de nuvem de pontos. Esta etapa é dividida em duas partes: 1) Detecta e extrai pontos correspondentes no par de imagens RGB empregando o algoritmo SIFT. Remove falsas correspondências através da Matriz Fundamental e do RANSAC. Gera as nuvens de pontos 3D para cada par de imagens processadas. Normaliza as imagens

IR em relação as imagens RGB e associa os pontos visuais correspondentes com as nuvens de pontos 3D por meio de uma geometria epipolar. E finalmente, estima os parâmetros de transformação do registro grosseiro usando um modelo matemático rigoroso e o MMQ; 2) Divide a imagem IR de referência em células igualmente espaçadas. Calcula o centroide de cada célula e o ponto 3D usando as Equações (4)-(6) e estabelece as correspondências entre o centroide de cada uma das células com os vizinhos mais próximos na nuvem de pesquisa usando o método *k-d tree*. Aplica o ICP e refina os parâmetros de transformação do registro grosseiro. Finalmente, é aplicado o ajustamento por pares de nuvens de pontos independentes.

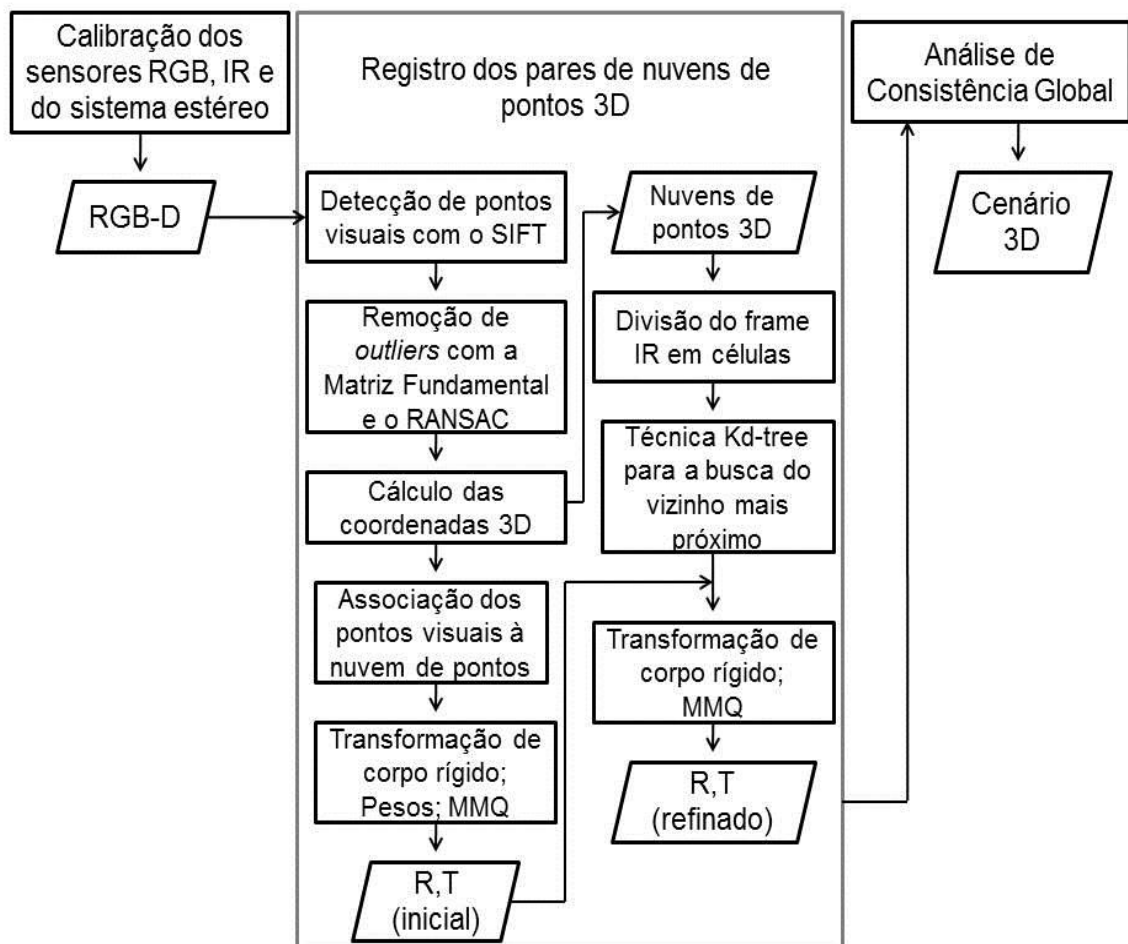


FIGURA 10 - DIAGRAMA DO MÉTODO PROPOSTO

FONTE: O autor (2015)

A seguir, serão apresentados cada etapa do método proposto descrito no diagrama da Figura 10.

3.4.1. Calibração do sistema

Esta etapa consiste em determinar os POI correspondentes aos sensores RGB e IR, e, os parâmetros de montagem do dispositivo Kinect. Para determinar os POI são empregadas as Equações (16)-(19) através do método de Zhang (2000). De acordo com Zhang (2000) primeiramente é obtida uma solução aplicando as Equações (15-19). Para refinar a solução obtida, deve ser empregada uma estimativa por máxima verossimilhança. Assumindo que p'_i esteja isento de erros sistemáticos, cujas coordenadas possuem a mesma incerteza na medida, então, a matriz H pode ser obtida da seguinte forma, a saber (MORE, 1997):

$$\min_H \sum_{i=1}^n \|p_i - p'_i\|^2 \quad (40)$$

sendo, $p'_i = \frac{1}{\bar{h}_3^T \cdot P_i} \begin{bmatrix} \bar{h}_1^T \cdot P_i \\ \bar{h}_2^T \cdot P_i \end{bmatrix}$, \bar{h}_i^T as linhas de índice i da matriz homográfica H , p_i as coordenadas do ponto no espaço-imagem e P_i as coordenadas do ponto correspondente no espaço-objeto.

Segundo Zhang (2000), uma forma natural de determinar simultaneamente todos os parâmetros envolvidos no problema de calibração pode ser feita pelo método de Levenberg-Maquardt, como segue:

$$\min \sum_{i=1}^n \sum_{j=1}^m \|p_{i,j} - p'(A, k_1, k_2, k_3, P_1, P_2, R_i, t_i, P_j)\|^2 \quad (41)$$

sendo, m o número de pontos na imagem, n o número de imagens, $p'(A, k_1, k_2, k_3, P_1, P_2, R_i, t_i, P_j)$ representa a projeção do ponto P_j na imagem i , R_i a matriz das rotações e t_i o vetor das translações. Vale lembrar que p' é função das coordenadas corrigidas das distorções do sistema de lentes como apresentada na Equação (8). Esse método de Levenberg-Maquardt é semelhante ao método combinado do MMQ, embora a diferença resida na forma como é invertida a matriz das derivadas parciais. Portanto, haja vista que é um problema não-linear, os

valores iniciais podem ser obtidos fazendo $H' = [h_1^T, h_2^T, h_3^T]^T$, então a Equação (9) pode ser reescrita como (ZHANG, 2000):

$$\begin{bmatrix} P^T & 0^T & -x_0 P^T \\ 0^T & P^T & -y_0 P^T \end{bmatrix} \cdot H' = 0 \quad (42)$$

Reescrevendo a Equação (42) da forma matricial $E \cdot H' = 0$, a solução pode ser obtida através do auto vetor de $E^T E$ associado ao menor auto-valor. Maiores detalhes ver Zhang (2000).

3.4.1.1. Normalização dos Valores de Profundidade

Esta etapa do método visa estimar os parâmetros, b e Z_0 usando o método empregado por Khoshelham e Elberink (2012). Considerando que esses parâmetros estão envolvidos em uma relação linear entre valores de paralaxe e de profundidade, na prática, a Equação (4) pode ser reescrita em função dos valores, c_2 e c_3 da normalização linear e da paralaxe normalizada, como segue:

$$Z_k = 1/c_2 \cdot d' + c_3 \quad (43)$$

sendo, c_2 e c_3 a interseção e a inclinação da linha reta, respectivamente.

Neste trabalho, é proposto o emprego de um interferômetro a LASER da marca Hewlett Packard modelo 5508A para medir os deslocamentos entre o sensor e uma placa de invar com o objetivo de poder calcular as grandezas de c_2 e c_3 . O Kinect foi posicionado de frente a uma placa de invar (ver Figura 11) e foram medidos com o interferômetro LASER 33 deslocamentos com 10 cm cada da plataforma contendo essa placa e o espelho reprojeto do LASER e, em cada intervalo, foi gerado com o Kinect um arquivo com as paralaxes. Assim, através das paralaxes normalizadas de um ponto na placa e do inverso das distâncias medidas com o interferômetro foi realizado o ajuste por mínimos quadrados dessas medidas

resultando nos parâmetros, c_2 e c_3 . A partir da obtenção desses valores é possível calcular os valores de profundidade de acordo com as paralaxes normalizadas observadas, como mostra a Equação (43).



FIGURA 11- EXPERIMENTO REALIZADO COM O INTERFERÔMETRO E COM O KINECT PARA A NORMALIZAÇÃO DOS VALORES DE PROFUNDIDADE. (a) IMAGEM RGB; (b) IMAGEM DE PROFUNDIDADE

FONTE: O autor (2015)

Vale ressaltar que o interferômetro a LASER propicia medidas com precisão em torno 0,01 mm.

3.4.2. Registro de Pares de Nuvens de Pontos

Esta etapa do método consiste em determinar os parâmetros de transformação dos pares de nuvens de pontos 3D. Isto é feito da seguinte forma: os parâmetros de transformação são obtidos para cada par de nuvem de pontos 3D adquiridas nos instantes t e $t + 1$, $t + 1$ e $t + 2$, $t + k_1$ e $t + num$, sendo num o número total de *frames* (IR) adquiridos e $k_1 = num - 1$. Cada *frame* corresponde a uma nuvem de pontos 3D. Neste caso, os parâmetros de translação e rotação relativa entre as consecutivas nuvens de pontos são estimados indiretamente por meio do MMQ. Neste trabalho, esta etapa está dividida em duas partes, a saber:

1. Registro grosseiro: Consiste em determinar os parâmetros de transformação iniciais para melhorar a eficiência computacional e evitar divergência na solução do algoritmo ICP.
2. Refinamento dos parâmetros iniciais do registro: Consiste em refinar os parâmetros iniciais obtidos na etapa anterior empregando o ICP.

Também vale ressaltar que neste caso, mesmo que não seja possível determinar os parâmetros de transformação do registro grosseiro em função de regiões com pouca textura, o ICP poderá executar integralmente a tarefa de registro dos pares de nuvens de pontos.

3.4.2.1. Determinação dos Parâmetros do Registro Grosseiro

Primeiramente, é necessário transformar cada pixel no sistema referencial da imagem (coluna e linha) para um sistema com origem no centro da imagem e cuja orientação seja dextrogira (REISS, 2002). Essas transformações podem ser realizadas através de uma translação, uma reflexão no eixo y , e, ambos seguidos da aplicação de um fator de escala representado pelo tamanho do pixel, como segue:

$$x' = \left(Col - \frac{T_{Col} - 1}{2} \right) \cdot T_{p_x} \quad (44)$$

$$y' = \left(-Lin + \frac{T_{Lin} - 1}{2} \right) \cdot T_{p_y} \quad (45)$$

sendo, x' e y' as coordenadas referenciadas ao sistema com origem no centro da imagem, Col e Lin são coordenadas no sistema referencial da imagem, T_{Lin} e T_{Col} são o tamanho total da imagem em coluna e linha, e T_{p_x}, T_{p_y} são o tamanho do pixel em x e y (sistema cartesiano), respectivamente.

Em seguida, as coordenadas supracitadas são corrigidas dos erros sistemáticos como apresentado na Equação (8). A Figura 12 mostra a sequência dos passos a serem executados para a determinação dos parâmetros de transformação do registro grosseiro de cada par de nuvens de pontos.

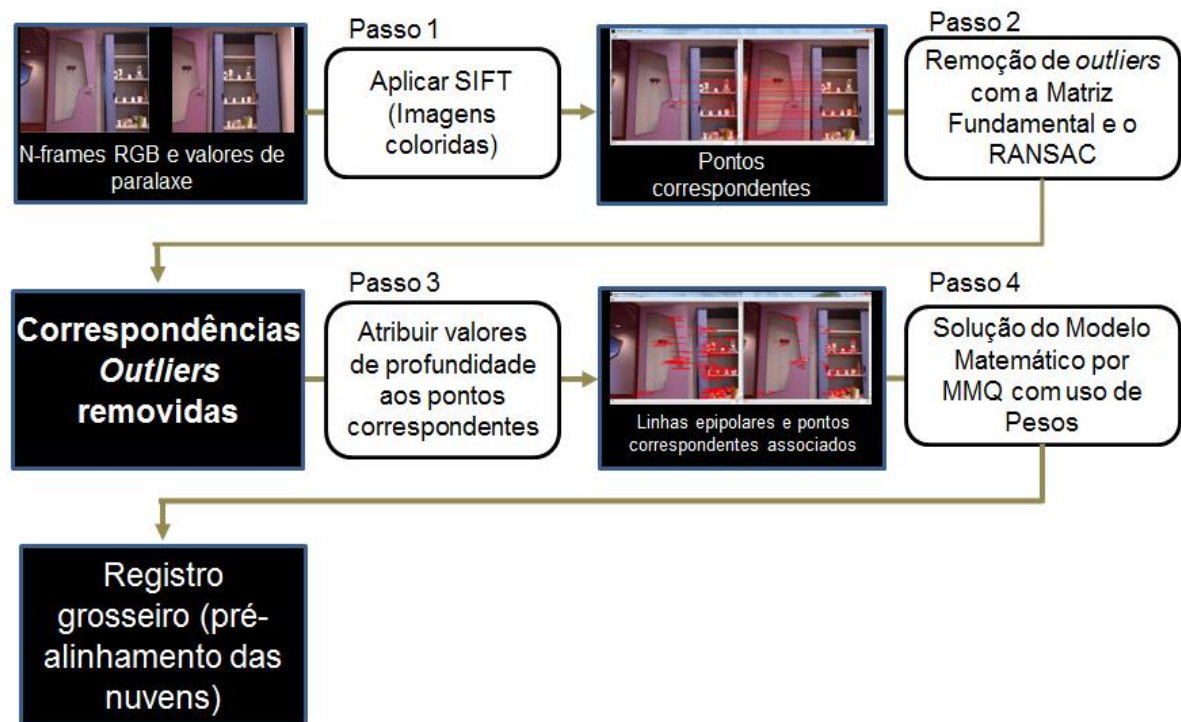


FIGURA 12 – SEQUÊNCIA DE PASSOS PARA A ETAPA DO REGISTRO GROSSEIRO

FONTE: O autor (2015)

Na Figura 12, dado um par de imagens RGB e os valores de paralaxe, primeiramente, o algoritmo emprega o SIFT para detectar pontos visuais na imagem RGB e uma geometria epipolar para encontrar as correspondências em potencial. Em seguida, são removidas do processo as falsas correspondências, e, posteriormente são atribuídos os valores de profundidade aos pontos correspondentes através de uma associação entre pontos na imagem RGB e na imagem de profundidade. Finalmente, é empregado o modelo de corpo rígido e o MMQ para a estimativa dos parâmetros de transformação do registro grosseiro com a inclusão de pesos baseados nas medidas tridimensionais.

3.4.2.1.1. Detecção de Pontos Visuais com o Algoritmo SIFT

Dado um par de imagens RGB, o objetivo é encontrar pontos visuais e calcular seus descritores de forma que possam ser estabelecidas suas correspondências. O algoritmo SIFT trabalha da seguinte forma, a saber:

1. Detecção de extremos no espaço escala através de cálculos de Diferença de Gaussianas (Equações 19-22);
2. Definição do local exato dos pontos encontrados através do ajustamento de uma função quadrática 3D (Equações 23 e 24);
3. Atribuição de uma ou mais orientações para cada localização dos pontos visuais baseadas em direções do gradiente local da imagem (Equações 25 e 26);
4. Composição dos descritores dos pontos visuais através de histogramas de orientação de regiões ao redor de cada ponto visual (Equação 27);

Portanto, para cada ponto visual detectado é atribuído um descritor invariante à iluminação, ao ponto de vista e à rotação. Neste trabalho, o descritor consiste de um vetor contendo os valores do histograma com 128 elementos determinados em função de uma janela de apoio com um total de 16x16 pixels. O resultado esperado é um conjunto de descritores que pode ser usado no estabelecimento automático de correspondências.

Finalmente, a correspondência entre os pontos de interesse é efetuada por intermédio do cálculo da medida de similaridade que descreve a relação entre os descritores em uma dada vizinhança.

3.4.2.1.2. Detecção e Remoção de Falsas Correspondências

Após serem encontradas as correspondências entre os pontos visuais, para cada par de imagens RGB é realizada a detecção remoção de falsas correspondências. Esta tarefa é feita em dois passos, a saber:

1. Geração da geometria epipolar com o cálculo da Matriz Fundamental através da associação entre os pontos visuais detectados (ver Equação 29) com o RANSAC e o MMQ;
2. Remoção de falsas correspondências usando um limiar de distância perpendicular da reta epipolar e o ponto correspondente na imagem de pesquisa;

Para calcular a matriz de transformação é necessário empregar a Equação (30, Subseção 2.3.2), que pode ser reescrita da seguinte forma, como segue:

$$A \cdot cf = 0 \quad (46)$$

Sendo A uma matriz de dimensão $nx9$, onde n representa o número de correspondências, a saber:

$$A = \begin{bmatrix} x'_{x1}x_{x1} & x'_{x1}y_{x1} & x'_{x1} & y'_{x1}x_{x1} & y'_{x1}y_{x1} & y'_{x1} & x_{x1} & y_{x1} & 1 \\ x'_{x2}x_{x2} & x'_{x2}y_{x2} & x'_{x2} & y'_{x2}x_{x2} & y'_{x2}y_{x2} & y'_{x2} & x_{x2} & y_{x2} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_{xn}x_{xn} & x'_{xn}y_{xn} & x'_{xn} & y'_{xn}x_{xn} & y'_{xn}y_{xn} & y'_{xn} & x_{xn} & y_{xn} & 1 \end{bmatrix} \quad (47)$$

Para solucionar cf é aplicado o MMQ. Entretanto, como há muitas correspondências entre pontos visuais, é necessário verificar quais devem ser usadas para gerar a melhor MF. Assim, são geradas diversas MF's, a partir de diferentes combinações, e então escolhida a ótima. Como esse processo é caro computacionalmente, o RANSAC é empregado. Conforme descrito na Seção 2.3.2, o RANSAC gera modelos a partir de um conjunto mínimo de dados. A partir destes modelos e utilizando o critério de parada com 30% de pontos remanescentes (*inliers*), ou seja, na iteração que se alcança 30% ou menos do número total de correspondências como *inliers*, é obtido o modelo mais representativo entre os avaliados.

Sendo assim, dado um ponto visual x na imagem de referência (esquerda) e x' na imagem de pesquisa (direita), é possível eliminar falsas correspondências desde que a distância δ perpendicular entre a reta epipolar originada do ponto na imagem de referência e o detectado na imagem de pesquisa seja menor que um determinado limiar l (ver Figura 13). Caso contrário, essa correspondência será considerada uma *outlier*. Portanto, para verificar se o ponto é um *outlier* é realizado um teste da distância através da equação:

$$l'_x = F \cdot x \quad (48)$$

sendo que, $F \cdot x$ define a linha epipolar l'_x do ponto x na imagem de referência e “.” representa o operador de multiplicação (ZHANG, 1998).

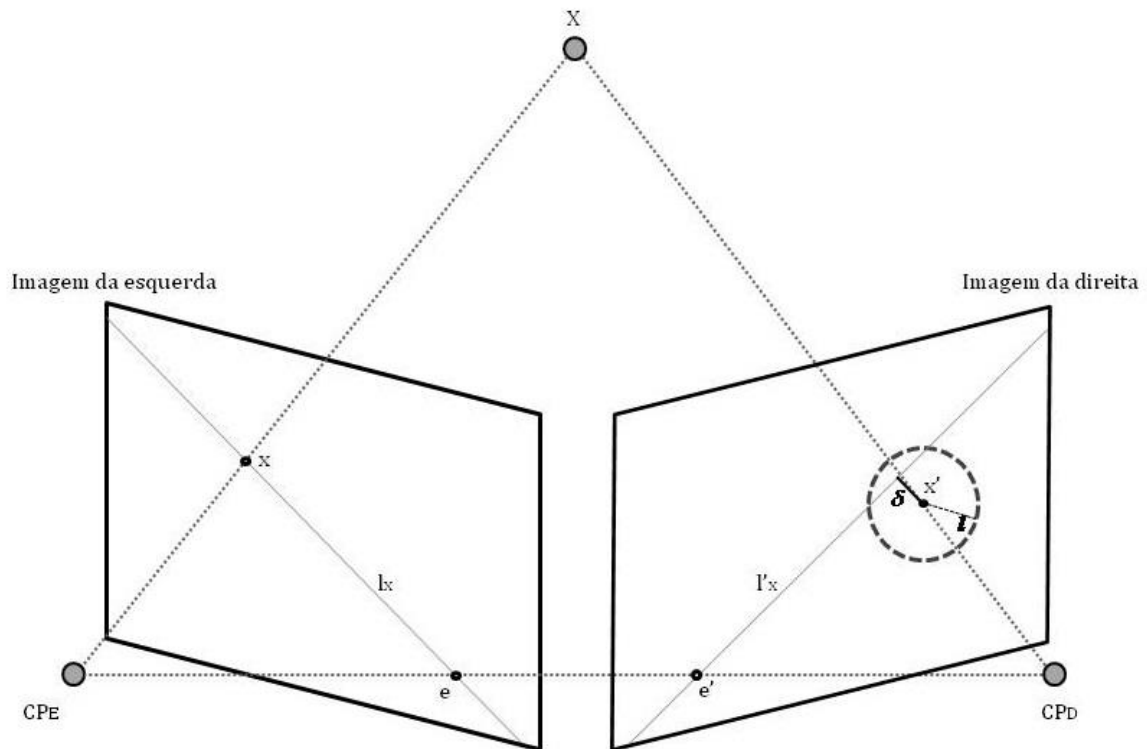


FIGURA 13- PROCESSO DE REMOÇÃO DE *OUTLIERS* COM O RANSAC

FONTE: ADAPTADO DE ZHANG (1998)

Na Figura 13, X é um ponto no espaço objeto, e e e' são os epipolos, CP_E e CP_D os respectivos centros perspectivos das imagens de referência (esquerda) e de pesquisa (direita). Outro critério que será usado neste trabalho para a remoção de falsas correspondências é o ângulo da reta que liga os pontos correspondentes, assim, se este ângulo for maior que 30 graus, esta correspondência deverá ser descartada.

3.4.2.1.3. Associação de Pontos Visuais na Nuvem de Pontos 3D

Cada par de pontos correspondentes detectado na imagem RGB deve ser associado ao seu ponto 3D correspondente na nuvem de pontos. Esta tarefa pode ser realizada a partir dos valores dos POI calibrados dos sensores RGB e IR, bem como dos parâmetros de montagem do dispositivo.

Como existe um desalinhamento angular entre as imagens RGB e IR, neste trabalho, é proposto uma normalização da imagem RGB. Essa é uma tarefa de

suma importância, tendo em vista que, a falta de correção dos desalinhamentos existentes entre os sensores RGB e IR provoca incertezas na associação direta dos pontos correspondentes da imagem RGB com a imagem de profundidade e, conseqüentemente, na determinação dos parâmetros de transformação. A normalização das imagens RGB consiste em 6 tarefas, a saber:

1. Primeiramente, a retificação da imagem é feita usando a Equação 49, como segue (MIKHAIL *et al.*, 2001):

$$\begin{aligned} x'_n &= -f_{IR} \frac{r_{11}x'_c + r_{12}y'_c - r_{13}f_{RGB}}{r_{31}x'_c + r_{32}y'_c - r_{33}f_{RGB}} \\ y'_n &= -f_{IR} \frac{r_{21}x'_c + r_{22}y'_c - r_{23}f_{RGB}}{r_{31}x'_c + r_{32}y'_c - r_{33}f_{RGB}} \end{aligned} \quad (49)$$

sendo, x'_n e y'_n as coordenadas normalizadas no sistema referencial com origem no centro da imagem IR, x'_c e y'_c as coordenadas no espaço-imagem corrigidas das distorções radial simétrica e descentrada do sistema de lentes da câmera IR, f_{RGB} e f_{IR} as distâncias focais calibradas dos sensores RGB e IR, respectivamente, r_{ij} os elementos da matriz de rotação composta pelos parâmetros angulares da montagem do dispositivo ($d\omega$, $d\phi$, $d\kappa$), para $i = j = 1, \dots, 3$;

2. Apesar de a imagem IR estar alinhada com a imagem RGB a diferença entre as origens dos sistemas referenciais não está compensada. Para tratar este problema uma linha epipolar é projetada na imagem IR. A projeção da linha é feita em função do valor de profundidade máximo e mínimo de alcance do dispositivo ($Z_{min} = 40cm$ e $Z_{max} = 7m$), do valor da linha base (b -calculado usando dx , dy , dz) e o valor de profundidade Z_k do ponto;

3. Para todos os pixels dentro do intervalo da linha epipolar são calculadas suas coordenadas 3D e reprojetaos para o sistema referencial digital da imagem RGB através da Equação (50), como segue:

$$\begin{aligned}
u &= f_{IR} \cdot \left(\frac{X_K}{Z_K} \right) - \left(\frac{x'_a}{T_{p_{xy}}} \right) \\
v &= f_{IR} \cdot \left(\frac{Y_K}{Z_K} \right) - \left(\frac{y'_a}{T_{p_{xy}}} \right)
\end{aligned} \tag{50}$$

sendo, u e v as coordenadas dos pixels reprojetados para a imagem IR normalizada, $T_{p_{xy}}$ referente ao sensor IR e x'_a, y'_a as coordenadas no sistema referencial do centro da imagem IR adicionadas dos erros sistemáticos (ponto principal e distorções das lentes da câmera IR).

4. Então, é verificado se o pixel transformado na 3ª tarefa corresponde ao pixel extraído pelo algoritmo SIFT na imagem RGB;

5. Para que suas coordenadas do ponto 3D correspondente sejam associadas ao ponto detectado na imagem RGB é necessário que a diferença entre eles seja nula;

6. As tarefas 1-5 são repetidas para todos os pontos detectados na imagem RGB.

Finalmente, é obtido um conjunto de pontos correspondentes associados à nuvem de pontos 3D. Desta forma, dado um conjunto de pontos 3D correspondentes, um modelo matemático rigoroso deve ser elaborado para determinar os parâmetros de transformação do registro grosseiro de cada par de nuvens de pontos 3D. Neste trabalho, o modelo empregado minimiza o somatório dos erros (E), como segue:

$$E = \sum_{m=1}^n w_{mp} \left\| X_{m,p-1} - \left(R \cdot X_{m,p} + \vec{T} \right) \right\| \tag{51}$$

sendo, $X_{m,p}$ e $X_{m,p-1}$ as coordenadas 3D correspondentes na nuvem de pontos M e P , respectivamente, w_{mp} o peso associado ao par de pontos correspondentes m e p , \vec{T} o vetor dos parâmetros de translação (t_x, t_y, t_z) e R a matriz de rotação em função dos ângulos de Euler (ϖ, ϕ, κ) .

Como a Equação (51) é não-linear deve ser aplicado um ajustamento pelo MMQ com o modelo paramétrico com iteração como mostra a Seção 2.3.3. Os

pesos w_{mp} são encontrados em função da precisão teórica das coordenadas tridimensionais e formarão a matriz dos pesos no MMQ.

3.4.2.1.4. Ponderação das Coordenadas Tridimensionais

Os valores dos pesos são definidos para todo par de pontos correspondentes (*inliers*) baseados na precisão teórica das coordenadas 3D (X_k , Y_k e Z_k). Sabendo que as imagens de profundidade são adquiridas em uma taxa de 20 a 30 quadros por segundo, resultando em pequenas diferenças de rotação e translação entre as imagens consecutivas, pode-se aproximar as equações de observação, como segue:

$$v_i = \|X_{m,p-1} - X_{m,p}\| \quad (52)$$

Para Khoshelham *et al.* (2013), os pesos podem ser definidos da seguinte forma:

$$w_{mp} = \frac{c}{\sigma_{v_i}^2} = \frac{c}{\sigma_{X_{m,p-1}}^2 + \sigma_{X_{m,p}}^2} \quad (53)$$

sendo, σ_X^2 a variância de cada ponto X e c uma constante arbitrária.

Segundo Kholhelham e Elberink (2012), assumindo que na Equação (4) os parâmetros de calibração estejam determinados acuradamente e d' uma variável randômica com uma distribuição normal, é possível propagar a variância da medida de paralaxe para obter a variância da medida de profundidade como segue:

$$\sigma_{Z_k}^2 = \left(\frac{\partial Z_k}{\partial d} \right) \cdot \sigma_{d'}^2 \quad (54)$$

Fazendo as devidas manipulações matemáticas e a normalização $d = c_0 \cdot d' + c_1$, se chega à expressão para o desvio padrão da profundidade (KHOSHELHAM e ELBERINK, 2012):

$$\sigma_{Z_k} = \left(\frac{c_0}{f \cdot b} \right) \cdot Z_k^2 \cdot \sigma_{d'} \quad (55)$$

Portanto, a variância da profundidade $\sigma_{Z_k}^2$ tem a seguinte relação com a variância da paralaxe normalizada medida $\sigma_{d'}^2$ (KHOSHELHAM *et al.*, 2013):

$$\sigma_{Z_k}^2 = c_2^2 \cdot \sigma_{d'}^2 \cdot Z_k^4 \quad (56)$$

sendo, $\sigma_{d'}$ o desvio padrão da medida de paralaxe normalizada

Substituindo a Equação (56) na (53) obtém-se a ponderação da medida de profundidade, como segue (KHOSHELHAM *et al.*, 2013):

$$w_{Z_k} = \frac{\psi}{z_{m,p-1}^4 + z_{m,p}^4} \quad (57)$$

sendo, na prática $\psi = 1$.

Neste trabalho também é proposto a inclusão dos valores de peso para as coordenadas em X_k e Y_k . Usando o mesmo raciocínio realizado anteriormente, mas agora para as coordenadas X_k e Y_k (ver Eqs. 4 e 5), tem-se:

$$\sigma_{\vec{X}}^2 = \frac{\vec{x}_c^2}{f^2} \cdot \sigma_{Z_k}^2 + \frac{Z_k^2}{f^2} \cdot \sigma_{\vec{x}_c}^2 \quad (58)$$

sendo, $\sigma_{\vec{X}}^2$ as variâncias das coordenadas X_k e Y_k , \vec{x}_c as coordenadas x'_c e y'_c no espaço-imagem e $\sigma_{\vec{x}_c}^2$ suas variâncias.

Substituindo as Equações (5) e (56) na Equação (58) tem-se:

$$\sigma_{X_k}^2 = c_2^2 \cdot \sigma_d^2 \cdot x'_c \cdot Z_k^2 + \frac{\sigma_{x'_c}^2}{f^2} \cdot Z_k^2 \quad (59)$$

Da mesma forma, substituindo as equações (6) e (56) na equação (58) tem-se:

$$\sigma_{Y_k}^2 = c_2^2 \cdot \sigma_d^2 \cdot y'_c \cdot Z_k^2 + \frac{\sigma_{y'_c}^2}{f^2} \cdot Z_k^2 \quad (60)$$

Substituindo as equações (59) e (60) na equação (57) obtém-se a ponderação das medidas em X_k e Y_k , como segue:

$$w_{X_k} = \frac{\psi}{Z_{m,p-1}^2 \cdot (X_{k,m,p-1}^2 + \mu) + Z_{m,p}^2 \cdot (X_{k,m,p-1}^2 + \mu)} \quad (61)$$

$$w_{Y_k} = \frac{\psi}{Z_{m,p-1}^2 \cdot (Y_{k,m,p-1}^2 + \mu) + Z_{m,p}^2 \cdot (Y_{k,m,p-1}^2 + \mu)} \quad (62)$$

sendo, $\psi = c \cdot c_2^{-2} \cdot \sigma_d^{-2}$, embora na prática adotado como 1 e $\mu = c_2^{-2} \cdot f^{-2} \cdot \sigma_d^{-2} \cdot \sigma_{x'y'}^2$ constantes definidas pelos parâmetros c_2 e f , e, $\sigma_{x'y'}^2 = \sigma_{x'}^2 = \sigma_{y'}^2$.

Obtidos os valores dos pesos é feito o ajustamento por MMQ calculando os parâmetros do registro grosseiro. Em seguida, esses parâmetros de transformação são aplicados à nuvem de pontos de pesquisa esperando que o desalinhamento inicial entre as nuvens de pontos de referência e de pesquisa seja minimizado.

3.4.2.2. Refinamento dos Parâmetros de Transformação Iniciais

A Figura 14 mostra a sequência dos passos a serem executados para o refinamento dos parâmetros usando o ICP.

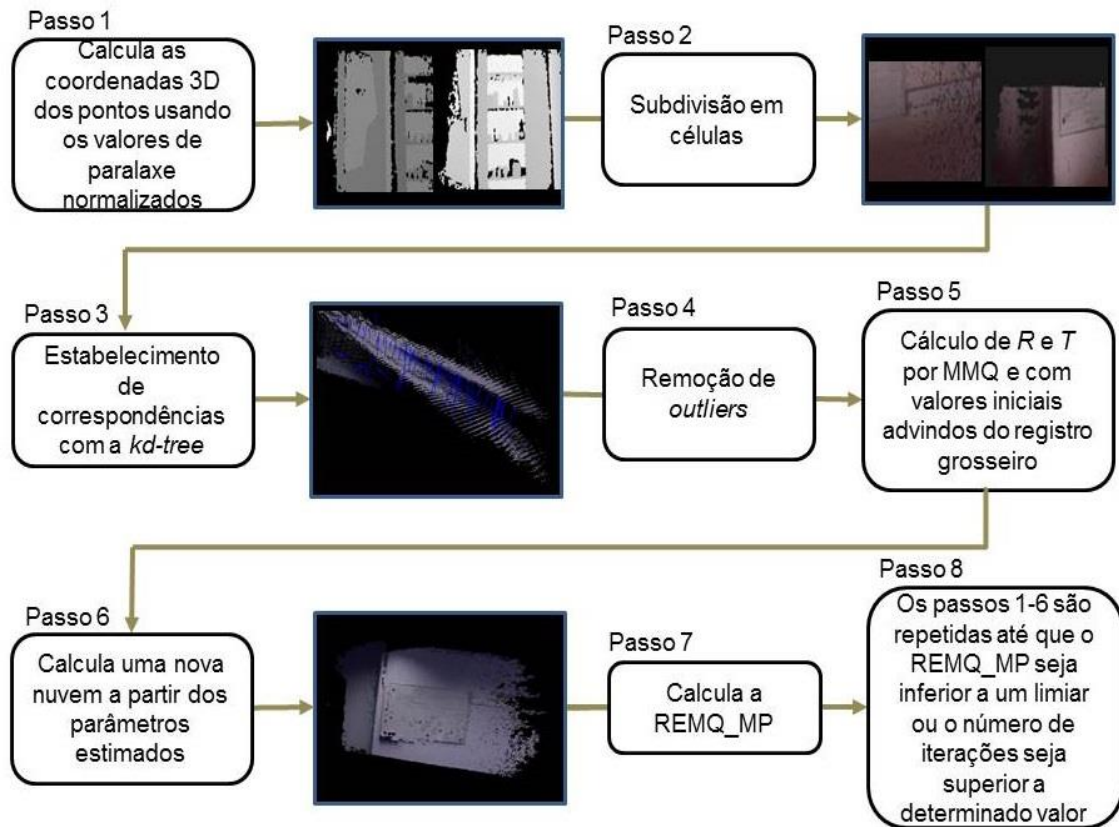


FIGURA 14 - SEQUÊNCIA DE PASSOS PARA A ETAPA DO REFINAMENTO DOS PARÂMETROS
FONTE: O autor (2015)

Na Figura 14, de posse dos parâmetros de transformação do registro grosseiro obtidos na etapa anterior é aplicado o algoritmo ICP com o objetivo de refinar esses parâmetros.

Neste trabalho é proposta uma estratégia adaptativa para o refinamento dos parâmetros de transformação do registro grosseiro. Neste caso, considere um par de imagens IR (imagem de referência e imagem de pesquisa) obtidas com o dispositivo Kinect. A imagem de referência é regularmente dividida em células de dimensão 10×10 centímetros. Para cada célula existe um conjunto de pixels, cujas coordenadas 3D são calculadas usando as Equações (4)-(6). Desta forma, é determinado o centroide (q) de cada célula, como segue:

$$q = \frac{1}{j} \sum m_j \quad (63)$$

sendo, m_j o j -ésimo ponto com coordenadas (X_k, Y_k, Z_k) contido na célula e j o número de pontos na célula.

Como resultado, tem-se um conjunto de centroides na nuvem de pontos de referência, cujas correspondências devem ser encontradas na nuvem de pontos de pesquisa. Vale destacar que para cada pixel na imagem de pesquisa também são determinadas suas coordenadas 3D através das Equações (4)-(6), sendo obtido um conjunto de p_i pontos de pesquisa.

Para acelerar o processo de busca pelo vizinho mais próximo foi empregada a estrutura *k-d tree*. Basicamente, o algoritmo realiza as seguintes tarefas, a saber: Dado um ponto centroide (q) e uma raiz da árvore de pesquisa (N) faz-se:

1. Para cada N divide-se todas as folhas da árvore de pesquisa ao longo de i usando um hiperplano que é perpendicular ao eixo correspondente;
2. Armazena-se dois sub-nós de cada N interno, um índice i que especifica a dimensão ao longo do qual N divide o espaço, e um escalar s que determina o ponto de divisão entre as duas sub-árvores;
3. Todos os pontos (p_i) mais próximos da nuvem de referência M são localizados;
4. O ponto mais próximo (p_i) de q que se intersecta em R (raio de distância) é o correspondente;
5. As tarefas 1-4 são repetidas até que seja pesquisado o último centroide;

Encontrado um conjunto de pontos correspondentes, a cada iteração, o modelo de corpo rígido 3D é empregado para refinar os parâmetros de transformação do registro grosseiro. O modelo de otimização minimiza o somatório do erro de reprojeção (E), explicitamente, para reduzir a distância euclidiana de todos os pontos correspondentes nas nuvens de pontos de referência e de pesquisa, como mostra a Equação (51) (sem ponderação no refinamento). Como descrito anteriormente, a Equação (51) não é linear e envolve 6 parâmetros a serem determinados pelo Método dos Mínimos Quadrados. Neste caso, deve ser aplicado o modelo paramétrico com iteração (ver Seção 2.3.3). A cada iteração, a distância entre o par de nuvens de pontos 3D (referência e pesquisa) é progressivamente reduzido, e uma nova nuvem de pontos 3D de pesquisa é obtida.

Neste trabalho, é proposto um método adaptativo para executar os passos 2 e 6. Uma vez que não se pode admitir correspondências exatas entre pontos em um par de nuvens de pontos 3D obtidas pelo Kinect ou LASER *scanner*, isto faz com que o modelo de otimização (Eq. 51) seja altamente sensível a ruídos, dada a forma como são atribuídos os pseudo-conjugados (processo padrão do ICP nas primeiras iterações). Outro fator que deteriora a qualidade dos dados Kinect é o fato de as imagens RGB e de profundidade não serem obtidas de forma sincronizada, fazendo com que a paralaxe calculada e suas coordenadas (x'_c e y'_c) correspondentes no espaço-imagem não sejam exatamente correspondentes. Isto afeta, principalmente, a qualidade posicional das coordenadas X_k e Y_k (Eqs. 5 e 6), que são calculadas em função das coordenadas no espaço-imagem e de Z_k . Para reduzir a probabilidade de utilizar *outliers* na etapa do registro de pares de nuvens de pontos, é proposto um método adaptativo para detecção e remoção de *outliers* (tarefa 4). Neste método, são definidos dois critérios: o da distância mínima entre os pontos; e o da variação angular da linha reta que liga dois pontos correspondentes em potencial.

O critério da distância mínima se adapta ao valor do REMQ_MP. Após cada iteração do algoritmo ICP, o REMQ_MP calculado é adotado como um novo valor de limiar. Somente pontos potencialmente correspondentes que sejam inferiores ao valor da REMQ_MP são considerados correspondentes. À medida que as diferenças angulares e lineares do registro diminuem, as correspondências entre os pontos tendem a ser melhores (apesar de nunca serem exatas). O critério da variação angular da linha reta que liga dois pontos correspondentes em potencial, também é usado para detectar e remover *outliers*. Neste caso, encontrado um par de pontos correspondentes é calculado o ângulo entre a linha reta, que liga os pontos correspondentes, e o eixo X do sistema referencial arbitrário do sensor.

Finalmente, o critério de parada do algoritmo também está em função do valor da REMQ_MP. O algoritmo termina o processamento caso a diferença da REMQ_MP calculada daquela iteração com a anterior seja menor que um determinado limiar. Isto significa que a estimativa dos dados está estável e o processamento deve ser finalizado. Este passo também reduz o tempo de processamento do algoritmo ICP.

A etapa do registro entre pares de nuvens de pontos deve ser aplicada enquanto o número de imagens for menor que $num + 1$.

3.4.3. Ajustamento por pares de nuvens de pontos independentes

Dado um conjunto de pares de nuvens de pontos 3D e seus parâmetros de transformação do registro devidamente refinados, a etapa de ajustamento por pares de nuvens de pontos 3D consiste em refinar e transformar o conjunto de parâmetros referenciados num sistema local em um sistema referencial global. Isto é feito através de um ajustamento simultâneo dos parâmetros do registro. Neste trabalho, é proposta uma simplificação da técnica apresentada por Lu e Millios (1997). A partir de um conjunto de nós e arestas (as relações espaciais entre as nuvens) é possível formar uma rede em que cada nó corresponde à posição de cada par de nuvens de pontos (s_i , para $i = 0, \dots$, número total de posições) em sua trajetória ($t_x^{s_{i,i+1}}, t_y^{s_{i,i+1}}, t_z^{s_{i,i+1}}$). A simplificação proposta neste trabalho é devido ao emprego direto dos parâmetros de translação juntamente com sua MVC, de forma simultânea. Neste caso, se deve ajustar um conjunto de observações através de um estimador ótimo usando as arestas combinadas em um circuito fechado para detecção de lugares revisitados. Diante à alta taxa de aquisição de dados pelo Kinect, os parâmetro de rotação foram ignoradas neste ajustamento global, uma vez que apresentam ângulos muito pequenos.

A formulação do algoritmo de estimação ótima, a partir do circuito de medidas conhecidas sobre $n + 1$ observações $X_0 = [t_x^{s_{10}}, t_y^{s_{10}}, t_z^{s_{10}}]^T$, $X_1 = [t_x^{s_{21}}, t_y^{s_{21}}, t_z^{s_{21}}]^T$, $X_n = [t_x^{s_{n,n+1}}, t_y^{s_{n,n+1}}, t_z^{s_{n,n+1}}]^T$ sugere que cada um deles represente um vetor 3×1 , e uma aresta L_{ij} entre duas observações X_i e X_j represente uma diferença mensurável. Neste trabalho, a equação de medida, referente à função L_{ij} de X_i e X_j , é dada na forma linear, como segue:

$$L_{ij} = X_i - X_j \quad (64)$$

Portanto, assumindo que há uma ligação entre todas as observações e para cada L_{ij} existe uma observação $\overline{L_{ij}}$ que é admitida portadora de uma distribuição gaussiana com média zero e covariância conhecida C_{ij} , a estimativa é ótima quando minimizada a seguinte distância de Mahalanobis, como segue (LU e MILIOS, 1997):

$$Q = \sum_{0 \leq i < j \leq n} (X_i - X_j - \bar{L}_{ij})^T C_{ij}^{-1} (X_i - X_j - \bar{L}_{ij}) \quad (65)$$

A partir da concatenação de todas as observações X_0, X_1, \dots, X_n (X), das diferenças $L_{ij} = X_i - X_j$ (L), e, inserindo uma matriz de incidência (A), formada por valores 1, -1, e 0, chega-se a equação linear de medida em sua forma matricial dada por (LU e MILIOS, 1997):

$$L = AX \quad (66)$$

Substituindo a Equação (66) na Equação (65) é possível representar função Q também em sua forma matricial como segue:

$$Q = (\bar{L} - AX)^T C^{-1} (\bar{L} - AX) \quad (67)$$

sendo, \bar{L} a concatenação de todas as observações \bar{L}_{ij} referentes à L_{ij} e C a covariância de \bar{L} . A solução para X que minimiza Q é dada por (LU e MILIOS, 1997):

$$X = (A^T C^{-1} A)^{-1} A^T C^{-1} \bar{L} \quad (68)$$

Para compor a matriz C , neste trabalho, é proposto o uso da MVC dos parâmetros de translação do registro obtidos na etapa de refinamento dos pares de nuvens de pontos. Assim, um ajustamento pelo MMQ é realizado e o novo conjunto de parâmetros é estimado simultaneamente baseado na equação descrita pela Equação (39). Como foi dito, as observações neste ajustamento são os valores de translação determinados na etapa do registro entre os pares de nuvens de pontos 3D.

A seguir serão apresentados os ensaios realizados e discutidos os resultados obtidos.

4. EXPERIMENTOS E RESULTADOS

Para avaliar o método proposto neste trabalho foram realizados 10 ensaios em cenários (A-J) que representam ambientes internos, tais como, quartos, salas, escritórios etc. O dispositivo Kinect usado neste trabalho foi programado para obter dados com uma frequência de 25 quadros por segundo. Para cada quadro é armazenado uma imagem RGB, uma imagem IR e um arquivo contendo os valores de paralaxe discretizados em 11 bits. Primeiramente foi realizada a calibração do Kinect bem como de seus sensores RGB e IR. Essa etapa pode ser dividida em duas tarefas: 1) calibração dos POI dos sensores RGB e IR embarcados no Kinect, bem como a determinação dos ângulos de desalinhamento e a distância entre as origens dos sistemas referenciais de cada sensor (POE); 2) a determinação dos parâmetros de normalização dos valores de profundidade c_2 e c_3 .

Os POI dos sensores RGB e IR foram obtidos por um processo de calibração de câmeras usando o método de Zhang. Esse método foi escolhido devido à facilidade que oferece no processo de calibração do Kinect. Para essa calibração foi utilizado o programa 'Kinect-stereo-calib' disponível gratuitamente pela MRPT. Portanto, foram tomadas, simultaneamente, 16 imagens RGB e IR sobre um tabuleiro de xadrez padrão em diferentes pontos de vista. A Figura 15 mostra o conjunto de imagens RGB e IR obtidas pelo dispositivo Kinect.

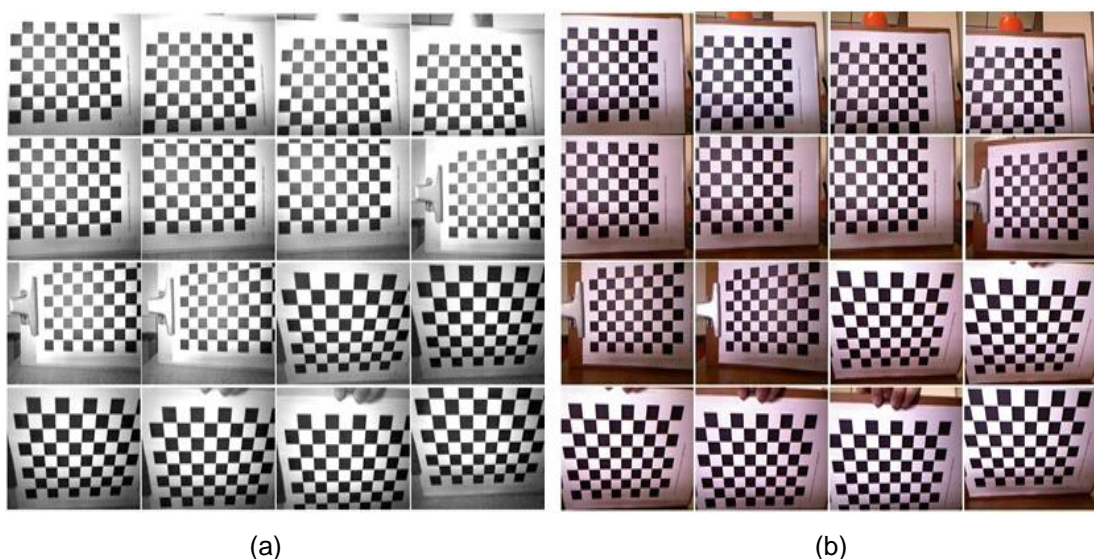


FIGURA 15 - CONJUNTO DE IMAGENS RGB E IR OBTIDAS PARA CALIBRAÇÃO DOS SENSORES. (A) IMAGENS IR; (B) IMAGENS RGB

FONTE: O Autor (2015)

Os POI de ambos os sensores e os POE podem ser encontrados na Tabela 1. Em relação à calibração estéreo, os parâmetros de translação e rotação são calculados da câmera da esquerda para a câmera da direita, ou seja, da IR para a RGB.

TABELA 1- PARÂMETROS DE ORIENTAÇÃO INTERIOR DOS SENSORES RGB E IR, SEUS DESVIOS PADRÃO E OS VALORES DOS POE.

Parâmetros	RGB	IR
f	$4,09 \pm 0,00036 \text{ mm}$	$4,55 \pm 0,00036 \text{ mm}$
$x_0 \text{ (offset)}$	$-0,0351 \pm 0,00008 \text{ mm}$	$-0,0273 \pm 0,00008 \text{ mm}$
$y_0 \text{ (offset)}$	$-0,1521 \pm 0,00008 \text{ mm}$	$-0,0897 \pm 0,00008 \text{ mm}$
k_1	$2,36 \times 10^{-1} \pm 8,0 \times 10^{-6} \text{ mm}^2$	$-7,76 \times 10^{-2} \pm 8,5 \times 10^{-6} \text{ mm}^2$
k_2	$-6,85 \times 10^{-1} \pm 7,5 \times 10^{-7} \text{ mm}^2$	$3,37 \times 10^{-1} \pm 8,0 \times 10^{-7} \text{ mm}^2$
k_3	$6,44 \times 10^{-3} \pm 8,0 \times 10^{-6} \text{ mm}^2$	$-4,68 \times 10^{-1} \pm 8,0 \times 10^{-7} \text{ mm}^2$
p_1	$1,24 \times 10^{-3} \pm 8,0 \times 10^{-6} \text{ mm}^2$	$-2,38 \times 10^{-3} \pm 8,0 \times 10^{-7} \text{ mm}^2$
p_2	$2,82 \times 10^{-3} \pm 8,0 \times 10^{-6} \text{ mm}^2$	$1,37 \times 10^{-3} \pm 8,0 \times 10^{-7} \text{ mm}^2$
$\Delta\omega(^{\circ})$	$\Delta\varphi(^{\circ})$	$\Delta\kappa(^{\circ})$
0,0240	-0,0001	-0,0011
		$\Delta T_x \text{ (mm)}$
		$\Delta T_y \text{ (mm)}$
		$\Delta T_z \text{ (mm)}$
		25,18
		$-4,1 \times 10^{-3}$
		$-9,3 \times 10^{-3}$

FONTE: O Autor (2015)

Como podem ser verificados na Tabela 1, os POE indicam que o centro do sensor RGB está localizado aproximadamente na mesma linha de base entre o sensor IR e o emissor LASER. Os POE encontrados são os esperados dada a integração dos sensores IR e RGB. A acurácia absoluta da calibração foi de 0,55 pixels para ambas as imagens, conforme os dados da REMQ do resíduo de reprojeção do ponto no espaço objeto.

Já os valores de c_2 (declive) e c_3 (intersecção da linha) foram obtidos usando o método proposto por Khoshelham e Elberink (2012). Ao contrário do que foi feito em Khoshelham e Elberink (2012), em que é usada uma fita métrica convencional para medir os deslocamentos do sensor frente a seu alvo, neste trabalho foi usado um interferômetro LASER disponível no Laboratório de Instrumentação Geodésica (LAIG). Para isto, foram obtidos os valores de paralaxe no centro da placa (ver Figura 12) para cada imagem de profundidade e plotados em relação ao inverso das distâncias do sensor para placa conforme se fazia os precisos deslocamentos com o interferômetro. A Figura 16 ilustra um gráfico que mostra a relação profundidade-paralaxe onde é possível perceber a relação linear expressada pela Equação (4) resultando nos valores para c_2 de $-3,07 \times 10^{-5}$ e para c_3 de $3,34 \times 10^{-2}$.

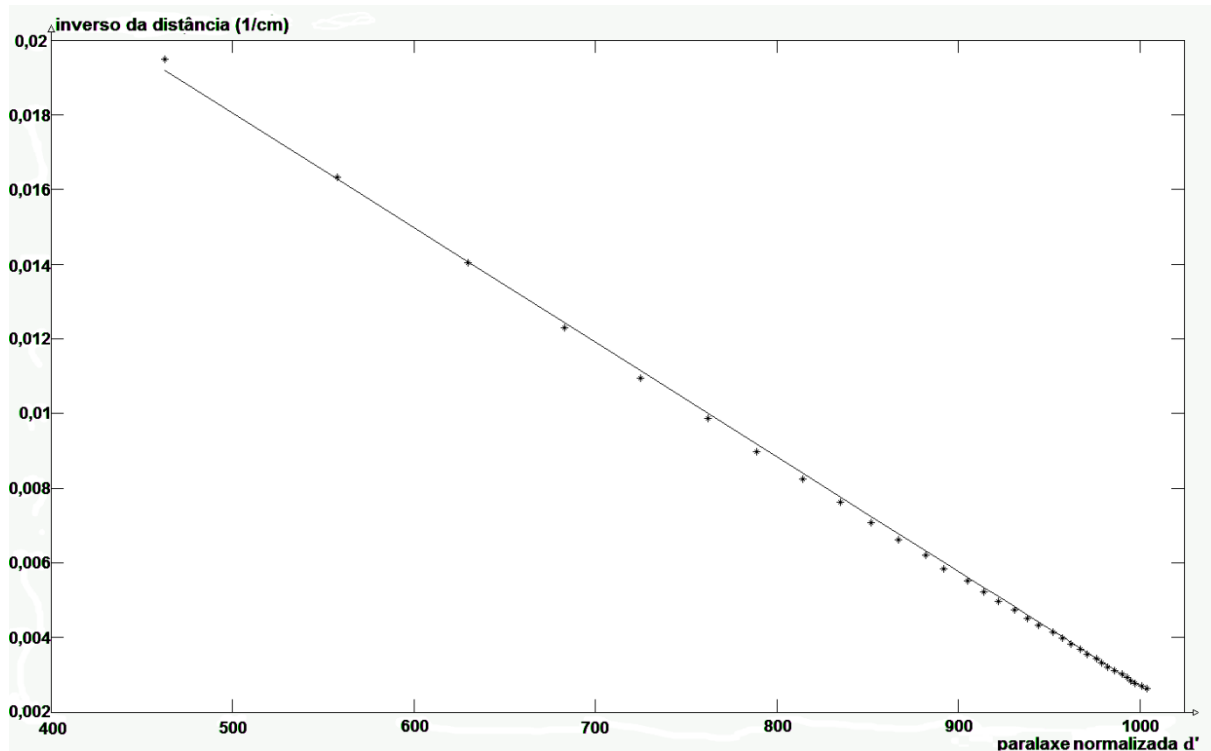


FIGURA 16 – RELAÇÃO LINEAR DA PARALAXE NORMALIZADA COM O INVERSO DA PROFUNDIDADE

FONTE: O Autor (2015)

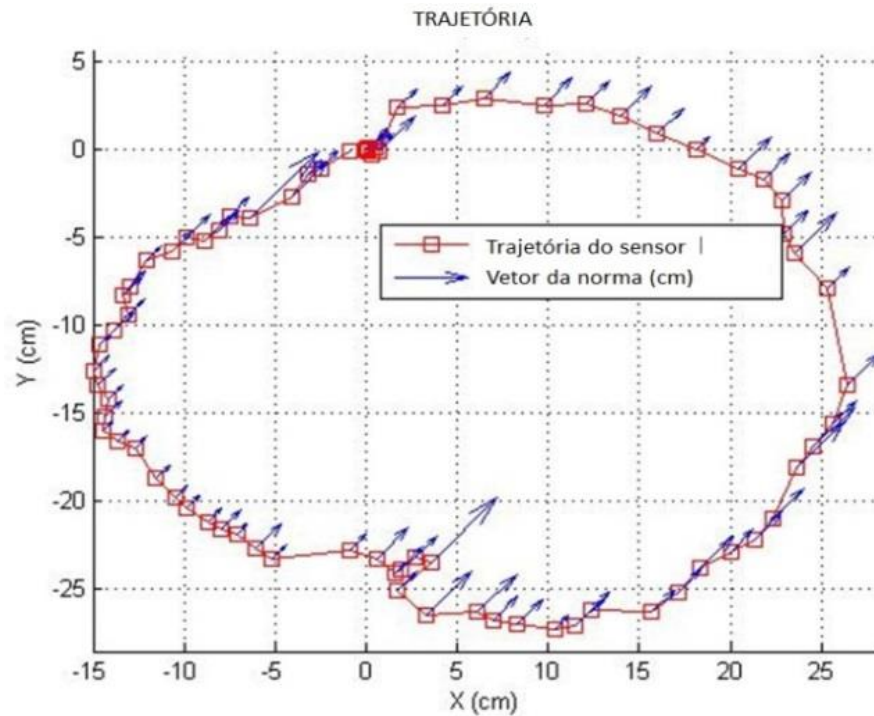
A partir da obtenção dos parâmetros da calibração foi possível conduzir os 10 ensaios propostos. Para este trabalho não serão usadas medidas verdadeiras das cenas, ou seja, pontos de controle com coordenadas absolutamente definidas, portanto, em cada experimento foi obtida uma trajetória formando uma volta completa apenas com o movimento circular das mãos para que, de certa forma, o método consiga registrar essas nuvens deixando a primeira e a última o mais coincidente possível. O Quadro 2 mostra o número de frames para cada experimento realizado para avaliar o método proposto.

Experimento	Número de frames
Cenário A	45
Cenário B	83
Cenário C	52
Cenário D	74
Cenário E	41
Cenário F	78
Cenário G	205
Cenário H	124
Cenário I	63
Cenário J	75

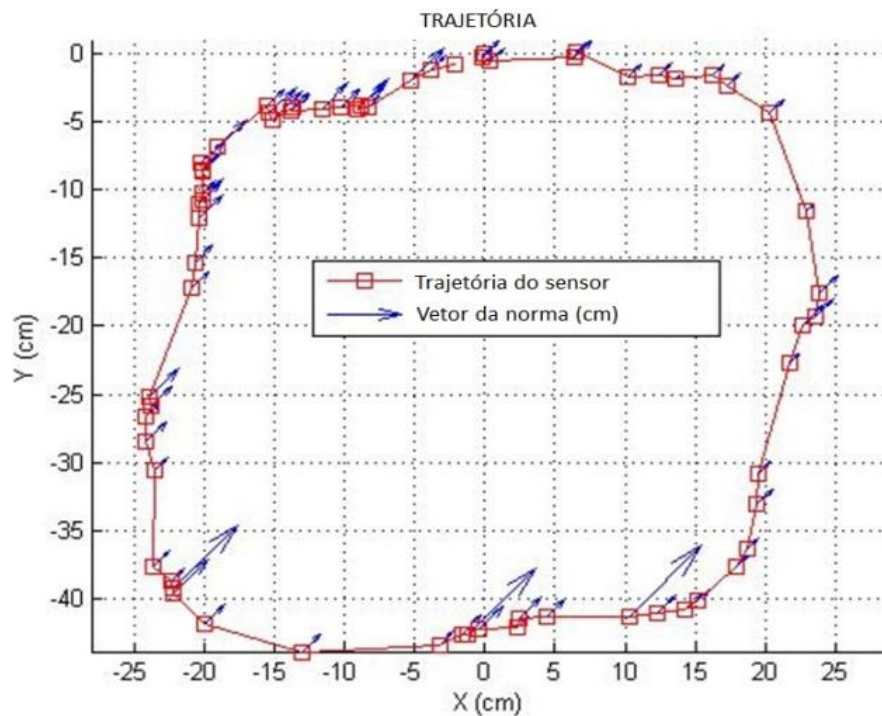
QUADRO 2 - EXPERIMENTOS REALIZADOS

FONTE: O autor (2015)

A Figura 17 mostra a trajetória do sensor depois da análise de consistência global (somente para os eixos X e Y) e o vetor da média das discrepâncias (relativo ao módulo) entre os pontos correspondentes de cada registro dos pares de nuvens de pontos (referência e pesquisa) obtidas com o método proposto para os cenários A, B e C.



(a)



(b)

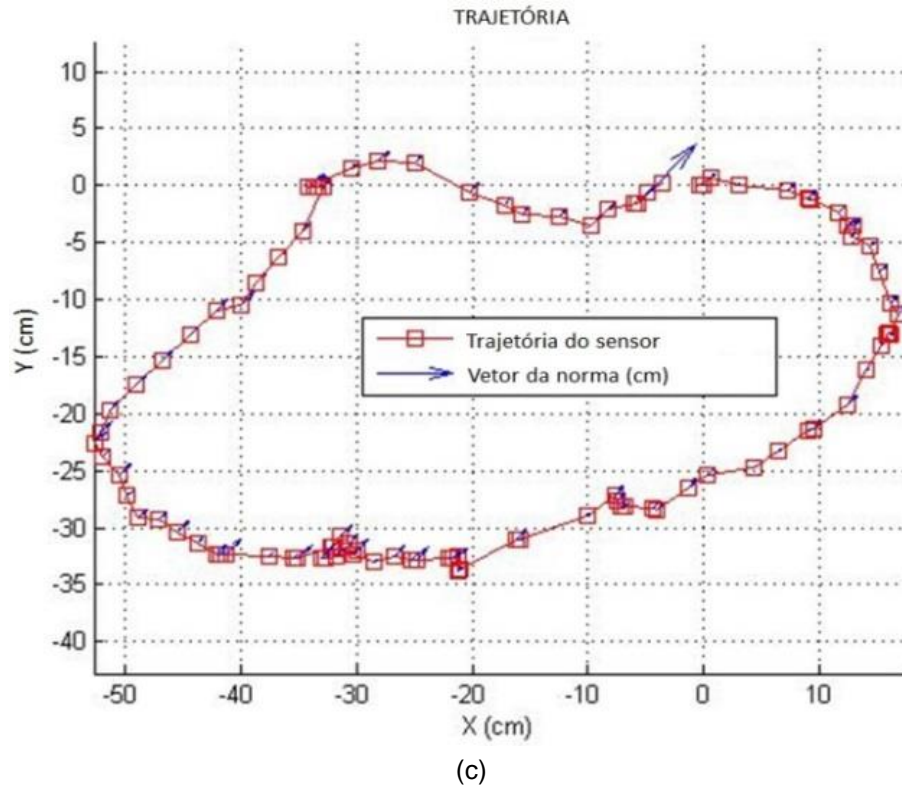
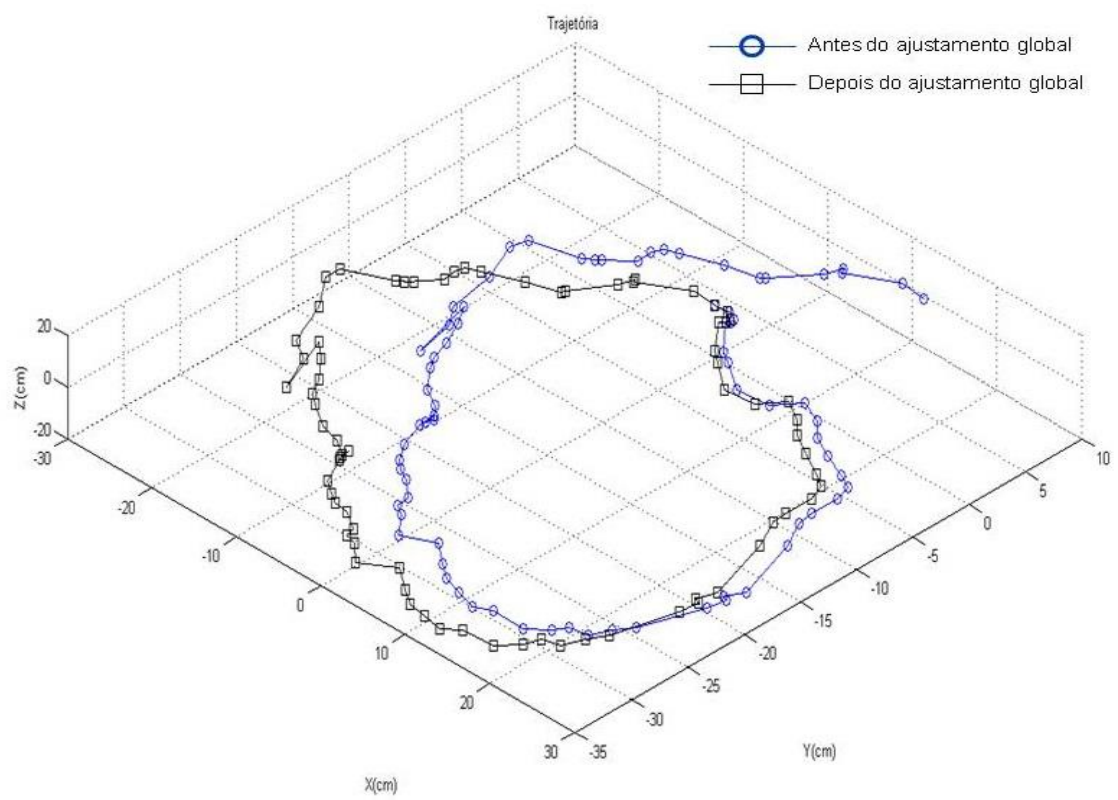
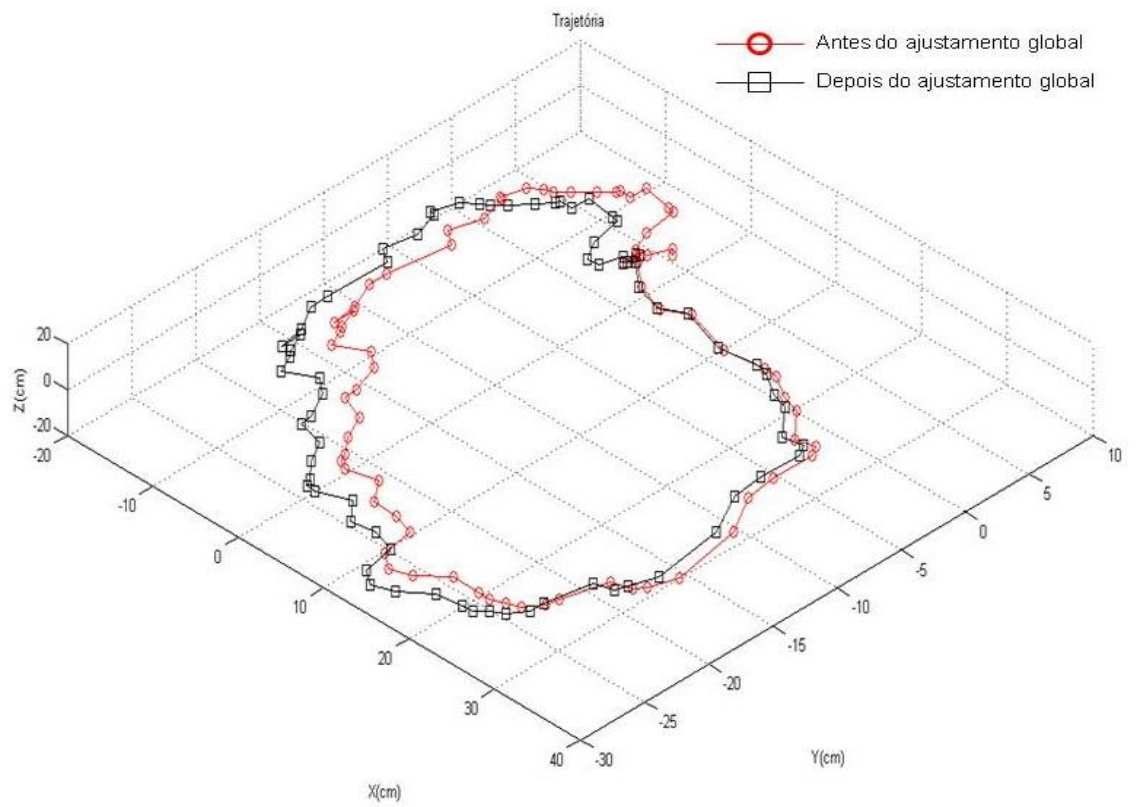


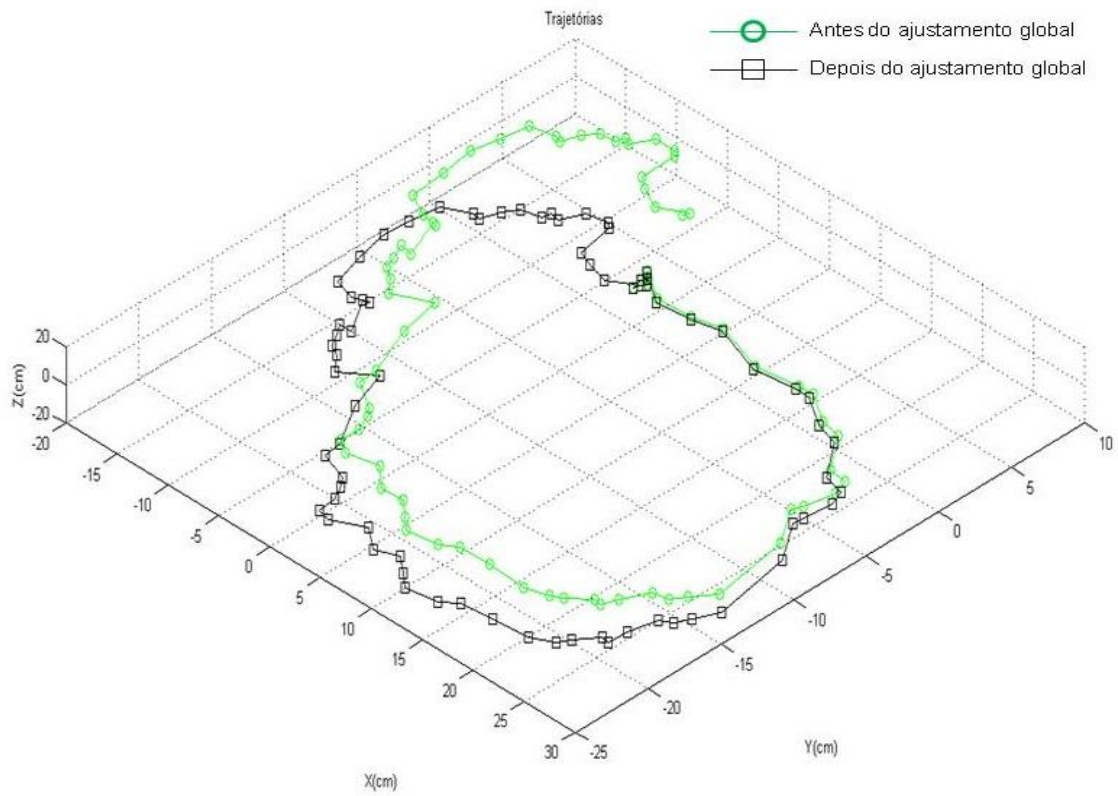
FIGURA 17 - TRAJETÓRIA DO SENSOR DEPOIS DA ANÁLISE DA CONSISTÊNCIA GLOBAL E O VETOR DAS DISCREPÂNCIAS MÉDIAS. (a) CENÁRIO A; (b) CENÁRIO B; (c) CENÁRIO C

FONTE: O autor (2015)

Na Figura 17, a trajetória do sensor, para os experimentos A, B e C, é obtida como o esperado. Antes da análise da consistência global foram encontrados valores da média das discrepâncias em torno de 10 cm e valores médios em torno de 3 cm após a análise da consistência global, sendo que as imagens foram coletadas a uma distância aproximada de 3 metros. No tocante as imagens, somente as magnitudes dos vetores são usadas para demonstrar os valores das médias das discrepâncias. Portanto, as inclinações dos vetores não demonstram nenhum comportamento sistemático da trajetória.

As Figuras 18, 19 e 20 mostram as trajetórias do sensor sem o uso de pesos (c, verde), com pesos em Z (a, vermelha), e, em XYZ (b, azul) para os cenários D, E e J, respectivamente. Nas mesmas figuras são plotadas as trajetórias (a, b e c, preta) ajustadas desses ensaios. Sendo que, a trajetória sem peso usou como observações os parâmetros de transformação do registro obtidos em um processamento sem ponderação, e assim sucessivamente para com peso em XYZ e somente em Z.

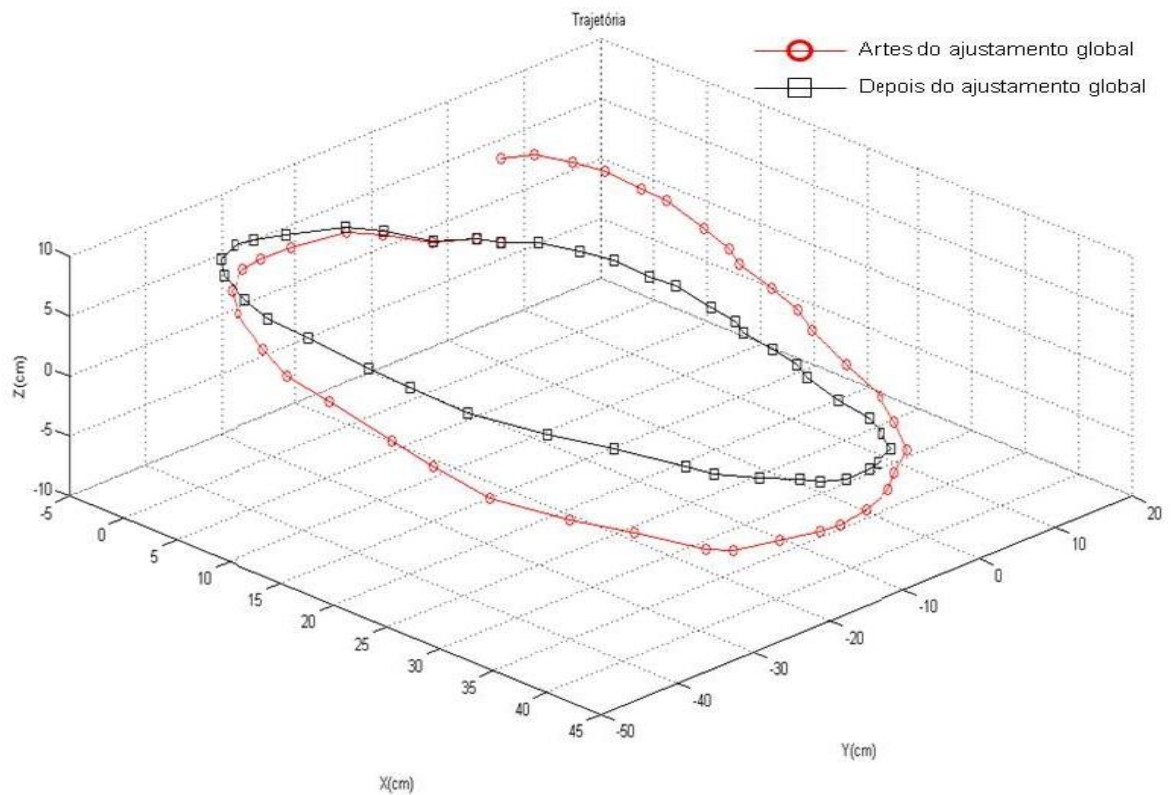




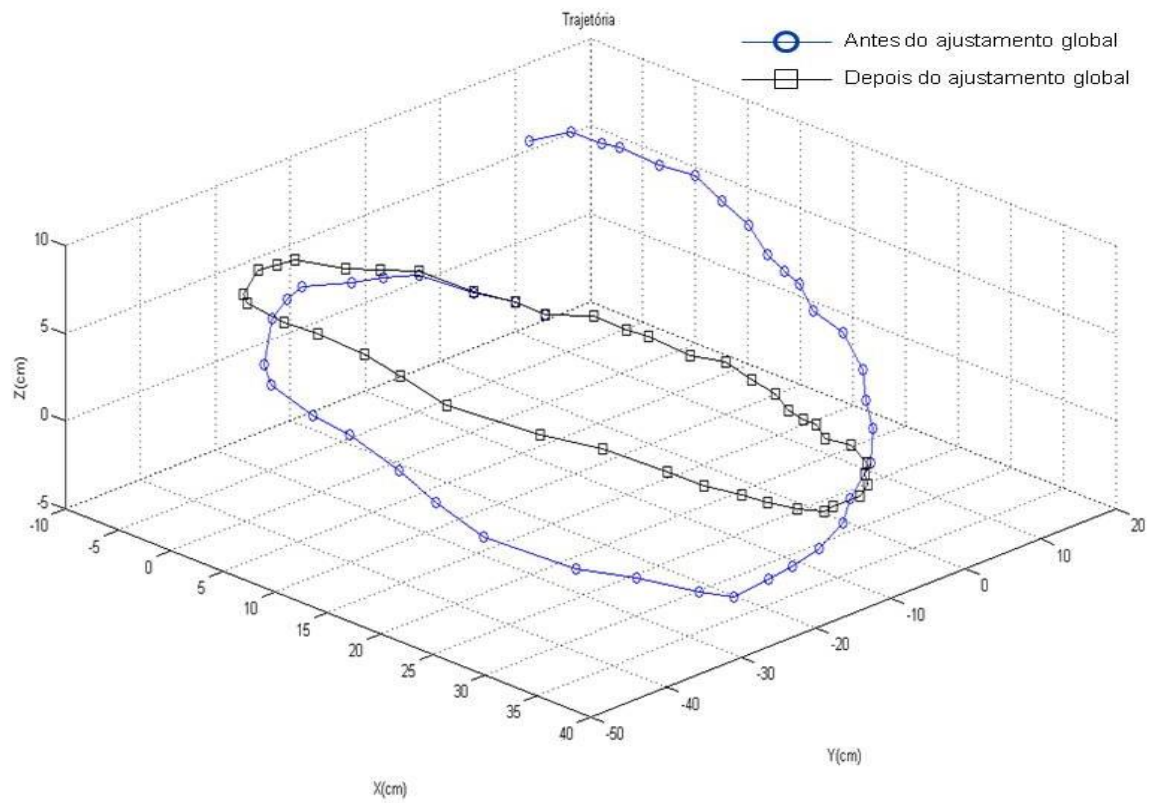
(c)

FIGURA 18 - TRAJETÓRIA DO SENSOR NO CENÁRIO D DEPOIS DA ANÁLISE DE CONSISTÊNCIA GLOBAL E ANTES COM PESO EM Z (a), EM XYZ (b) E SEM PESOS (c)

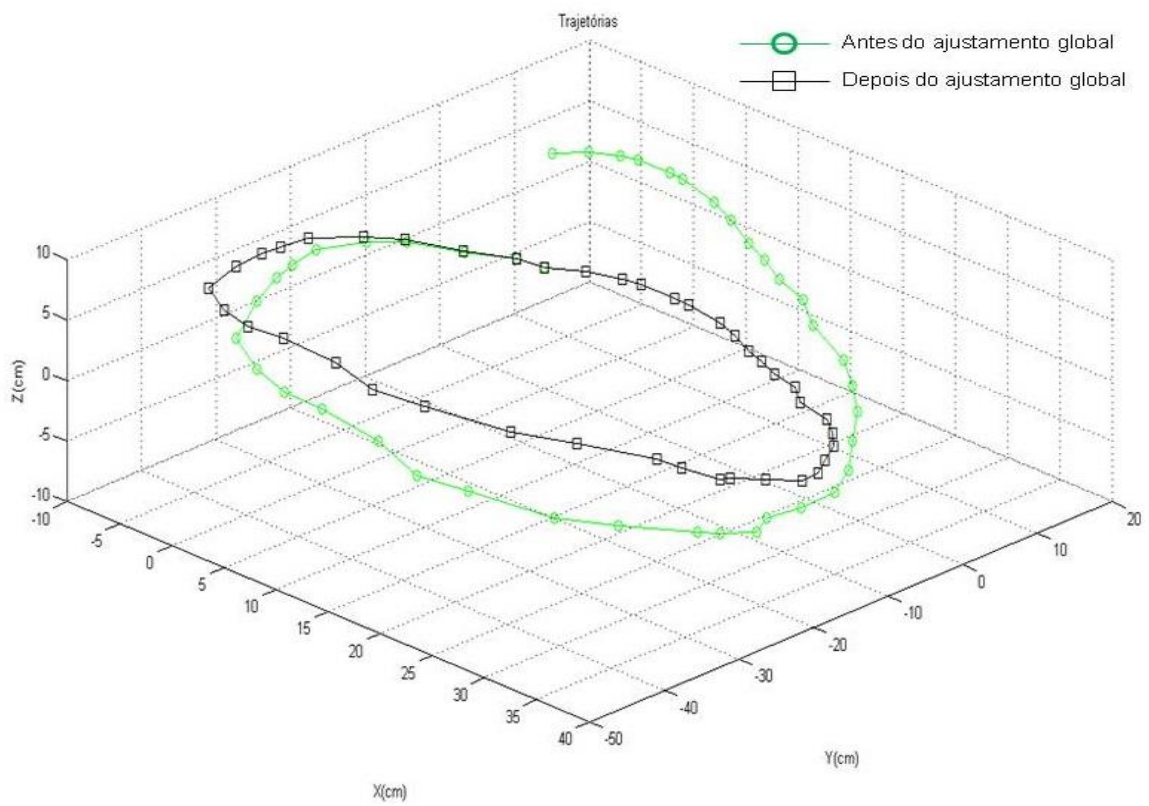
Fonte: O autor (2015)



(a)



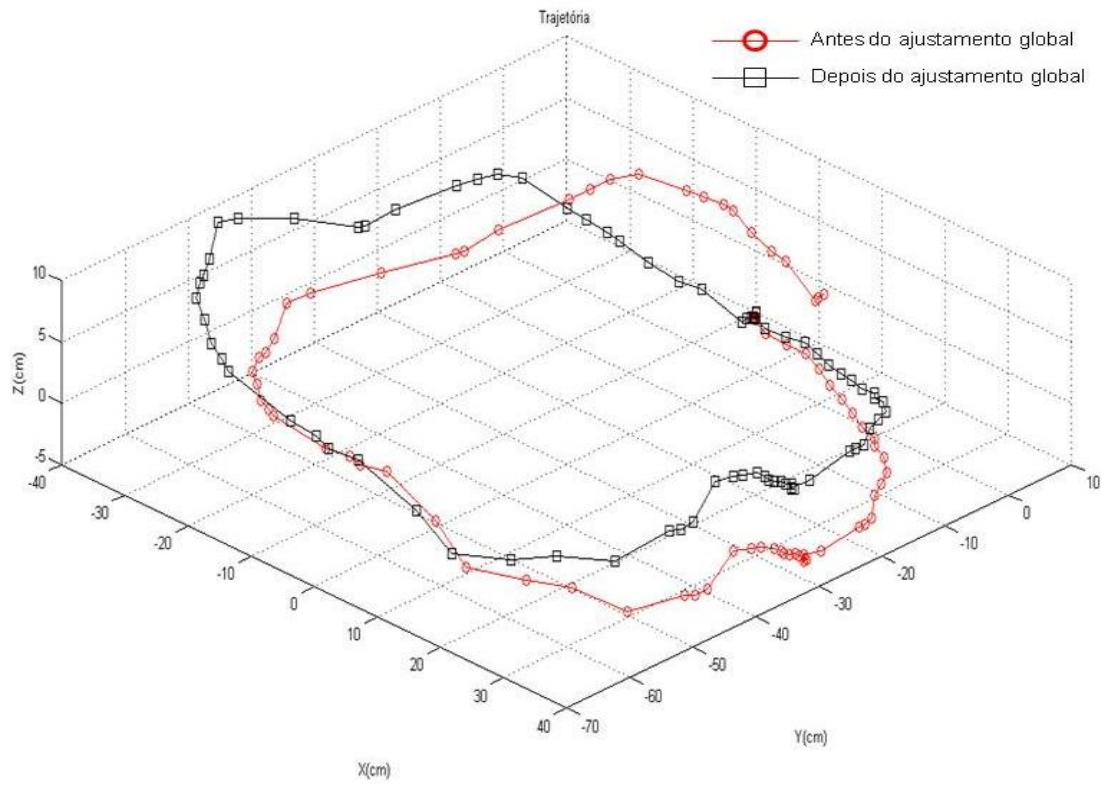
(b)



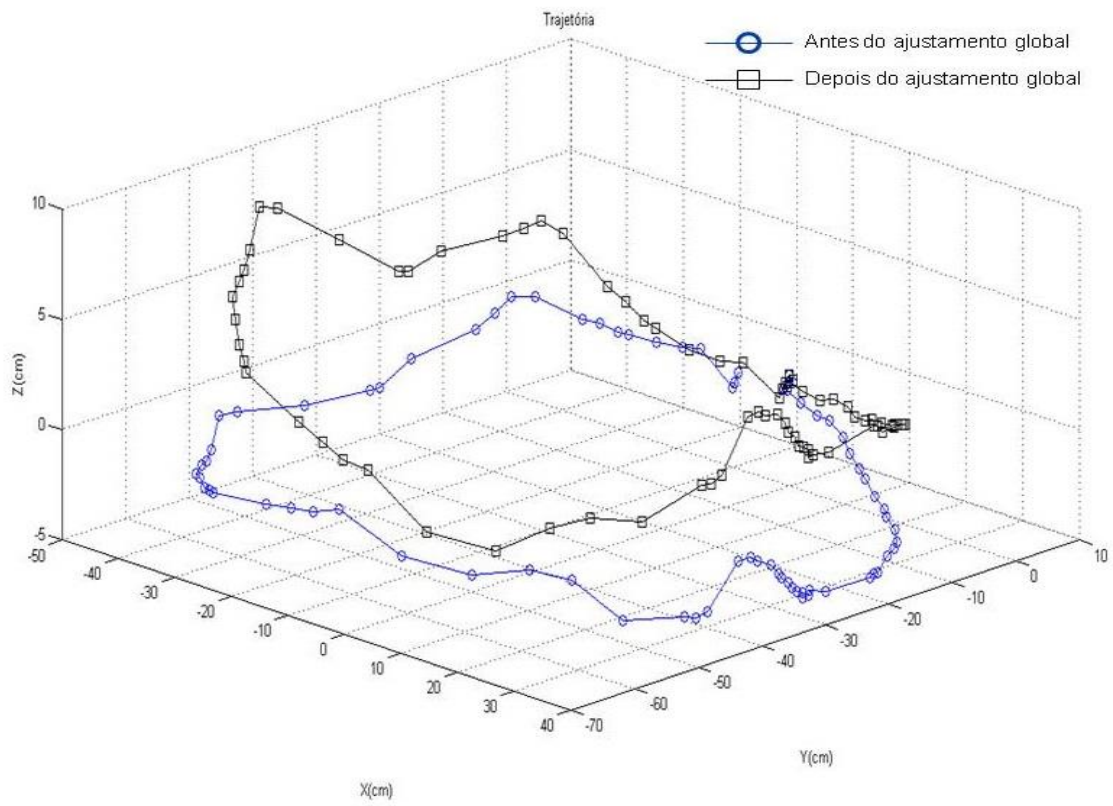
(c)

FIGURA 19 - TRAJETÓRIA DO SENSOR NO CENÁRIO E DEPOIS DA ANÁLISE DE CONSISTÊNCIA GLOBAL E ANTES COM PESO EM Z (a), EM XYZ (b) E SEM PESOS (c)

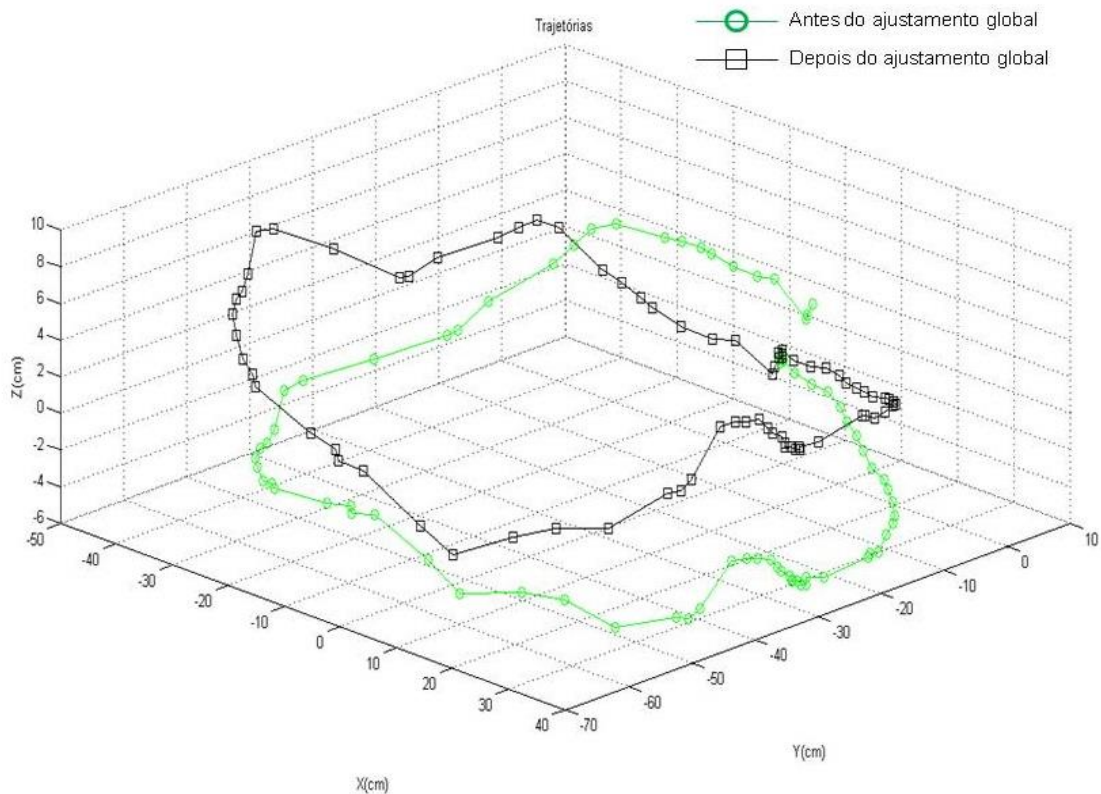
FONTE: O autor (2015)



(a)



(b)



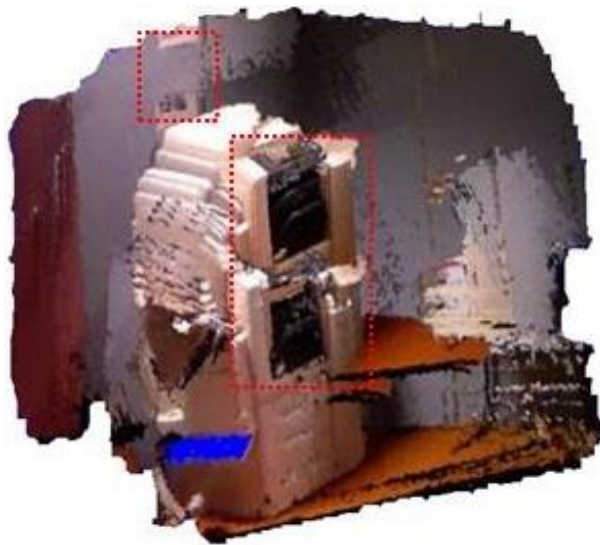
(c)

FIGURA 20 - TRAJETÓRIA DO SENSOR NO CENÁRIO J DEPOIS DA ANÁLISE DE CONSISTÊNCIA GLOBAL E ANTES COM PESO EM Z (a), EM XYZ (b) E SEM PESOS (c)

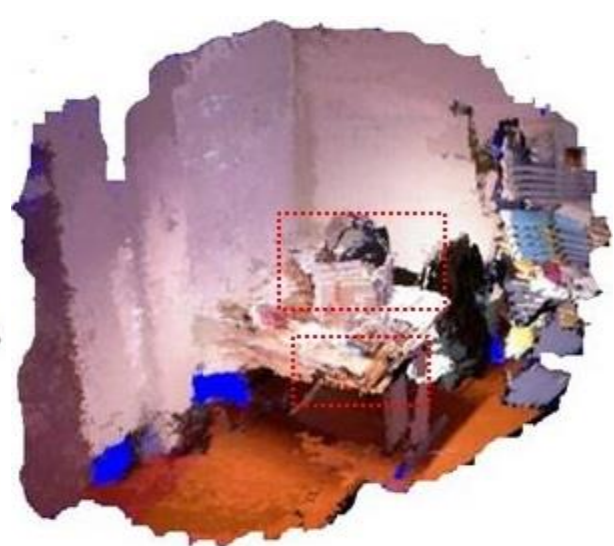
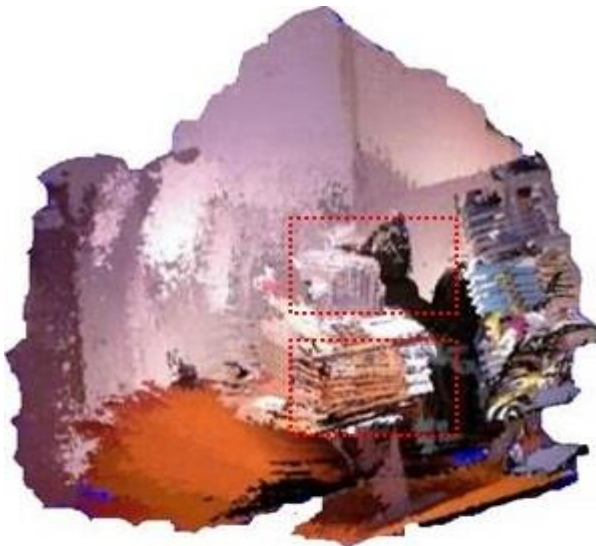
Fonte: O autor (2015)

Quanto ao comparativo do não uso de pesos, do uso em XYZ , e, em Z , no registro dos pares de nuvens de pontos 3D, as Figuras 18, 19 e 20 mostram que para os cenários D, E e J, respectivamente, a trajetória que mais se parece com a ajustada é aquela aplicando peso somente em Z (Figuras 18, 19 e 20; a). Além do mais, por meio desse experimento, mais uma vez é possível perceber como a propagação de erros age no registro dos pares de nuvens de pontos 3D quando comparado às trajetórias antes e depois do ajustamento global. Portanto, as figuras reafirmam a dependência de uma análise de consistência global consistente, que ajusta os parâmetros de transformação do registro fazendo com que o *loop* feche no final da trajetória, e, assim, gere nuvens de pontos visual e geometricamente melhores.

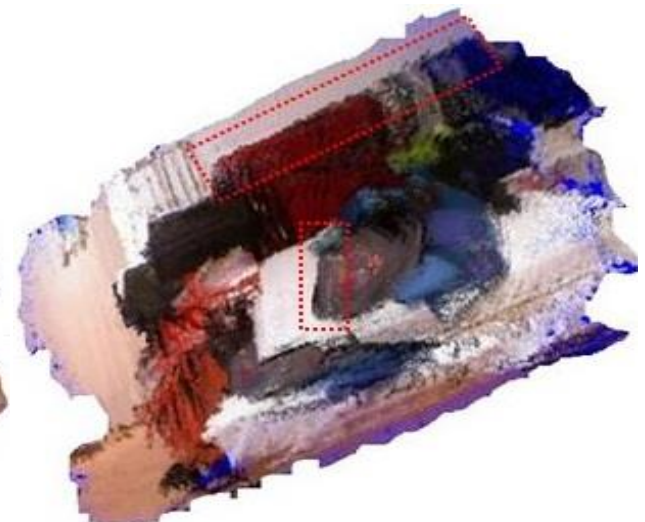
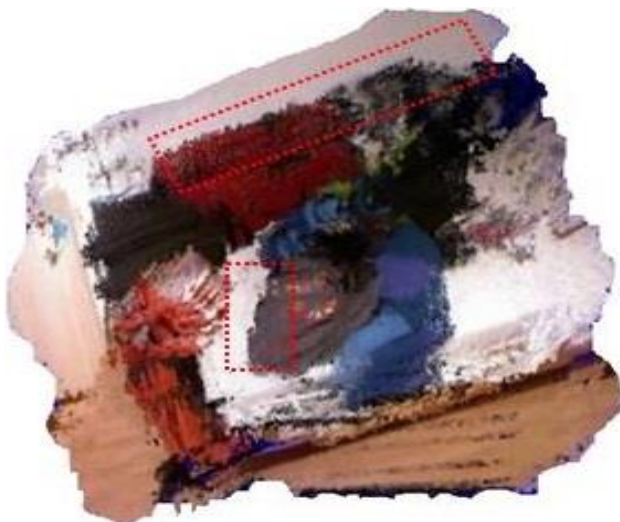
A Figura 21 mostra os ambientes internos tridimensionalmente mapeados antes e depois da análise de consistência global dos cenários A, B e C.



(a)



(b)



(c)

FIGURA 21 - NUVEM DE PONTOS 3D DE AMBIENTES INTERNOS MAPEADOS. (a) CENÁRIO A; (b) CENÁRIO B; (c) CENÁRIO C

FONTE: O autor (2015)

Na Figura 21a-c são apresentados os ambientes tridimensionalmente reconstruídos dos ensaios A, B e C através da aplicação dos parâmetros de transformação dos registros refinados antes e depois da análise de consistência global. Em todas elas são possíveis ver melhoras na qualidade visual dos ambientes depois da análise de consistência global (à direita) em comparação àqueles antes do ajustamento global (à esquerda). Como foi dito, isso acontece porque, antes do ajustamento global, a trajetória do sensor é afetada por erros acumulativos derivados da etapa de registro dos pares de nuvens de pontos 3D.

Em todas as imagens da Figura 21, estão destacadas, através de um retângulo pontilhado em vermelho, algumas regiões exemplo para demonstrar a melhora que ocorre depois do ajustamento global. Na Figura 21a, percebe-se uma fusão mais adequada da máquina de fotocópia e do folder anexo à parede após o ajustamento global (imagem da direita). Já na Figura 21b, claramente, algumas partes da feição da mesa são melhores fundidas após o ajustamento global. Por fim, na Figura 21c, os trechos lineares em destaque são melhor representados com o ajustamento global.

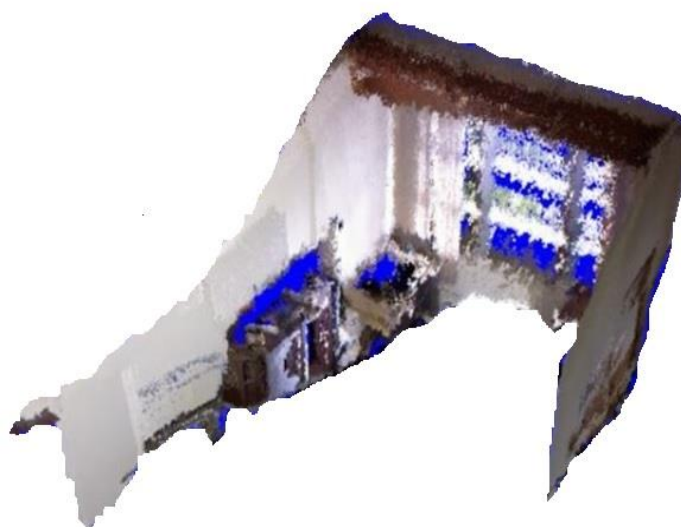
Já a Figura 22 mostra o ambiente interno tridimensionalmente mapeado depois da análise de consistência global dos cenários F, G e H.



(a)



(b)



(c)

FIGURA 22 - NUVEM DE PONTOS 3D DOS AMBIENTES INTERNOS. (a) CENÁRIO F; (b) CENÁRIO G; (c) CENÁRIO H

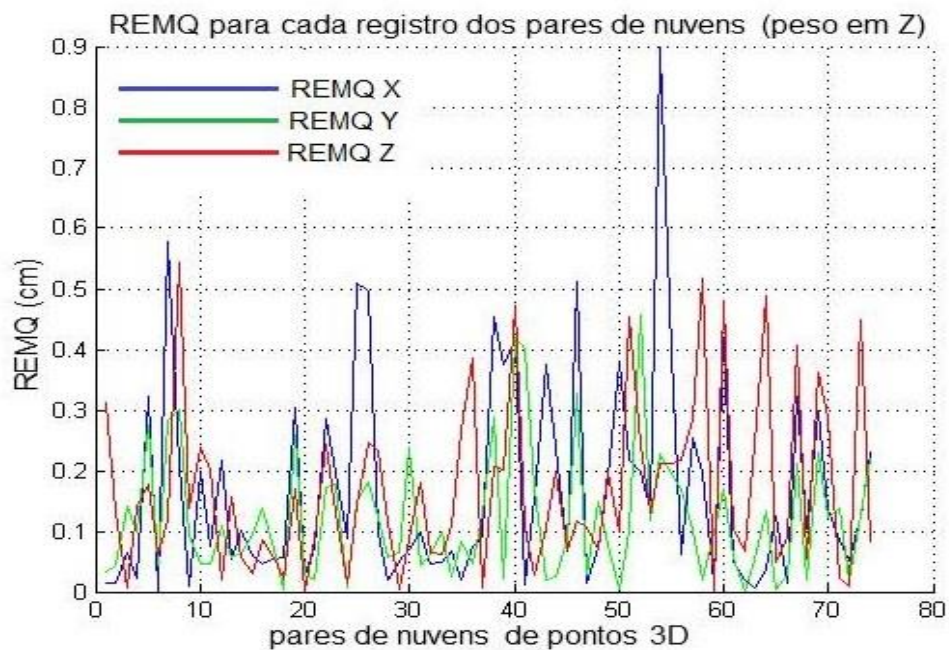
FONTE: O autor (2015)

A Figura 22a-c, mostra os ensaios F, G e H, respectivamente, e, os resultados apresentam uma ótima qualidade visual apesar da pouca texturização do ambiente, o que remete a uma das vantagens do método proposto que é manutenção da qualidade dos parâmetros de transformações advindos do registro mesmo com um número reduzido de *inliers* sendo detectados pelo SIFT. Mais adiante será apresentado um experimento que demonstra o comportamento do método frente a essas regiões.

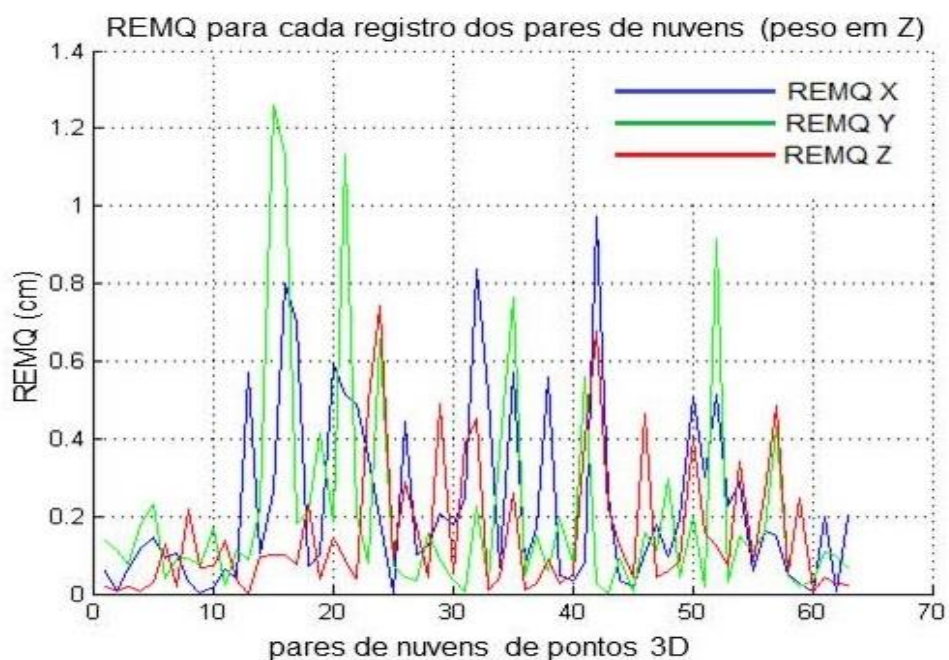
Para analisar quantitativamente os resultados obtidos foi calculada a REMQ das discrepâncias e a média das discrepâncias entre os pontos correspondentes encontrados na última iteração do refinamento com o algoritmo ICP (neste caso, os pontos são considerados correspondentes). A REMQ das discrepâncias dos experimentos D, I e E é exibida na Figura 23.

Os resultados mostram que o menor e o maior valor da REMQ das coordenadas, X , Y e Z calculados no experimento relativo ao cenário D foram zero e 0,9 cm para a coordenada em X , para a coordenada em Y foram obtidos zero e 0,46 cm, enquanto para a coordenada em Z foram encontrados valores de zero e 0,54 cm, respectivamente. Em relação ao cenário I foram obtidos valores da REMQ de zero e 0,89 cm para a coordenada em X , para a coordenada em Y foram determinados valores iguais a zero e 0,25 cm, enquanto para a coordenada em Z foram encontrados valores de zero e 0,75 cm, respectivamente. Para cenário E

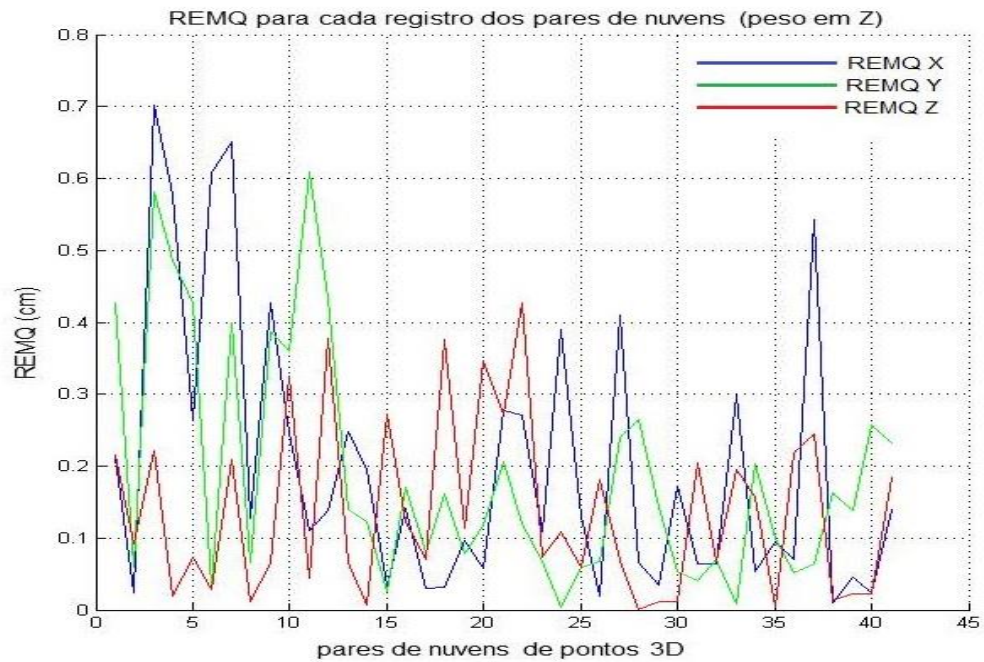
foram obtidos valores da REMQ de 0,03 e 0,7 cm para a coordenada em X , para a coordenada em Y foram determinados valores iguais a 0,01 e 0,60 cm, enquanto para a coordenada em Z foram encontrados valores de zero e 0,43 cm, respectivamente. Para todos os experimentos o valor da média das discrepâncias ficou próximo de zero, mostrando que nenhuma tendência significativa influenciou o modelo matemático.



(a)



(b)



(c)

FIGURA 23 - REMQ DA MÉDIA DAS DISCREPÂNCIAS CALCULADA EM CADA REGISTRO DE PARES DE NUVENS DE PONTOS. (a) CENÁRIO D; (b) CENÁRIO I; (c) CENÁRIO E.

FONTE: O autor (2015)

A Tabela 2 mostra os valores máximo e mínimo da média das discrepâncias para os experimentos A, B e C.

TABELA 2 - MÉDIA DAS DISCREPÂNCIAS APÓS A ANÁLISE DA CONSISTÊNCIA GLOBAL PARA OS CENÁRIOS A, B e C

Experimento	Média das discrepâncias (cm)
Cenário A	Min: 0,51 Max: 3,84
Cenário B	Min: 0,21 Max: 1,60
Cenário C	Min: 0,11 Max: 3,00

FONTE: O autor (2015)

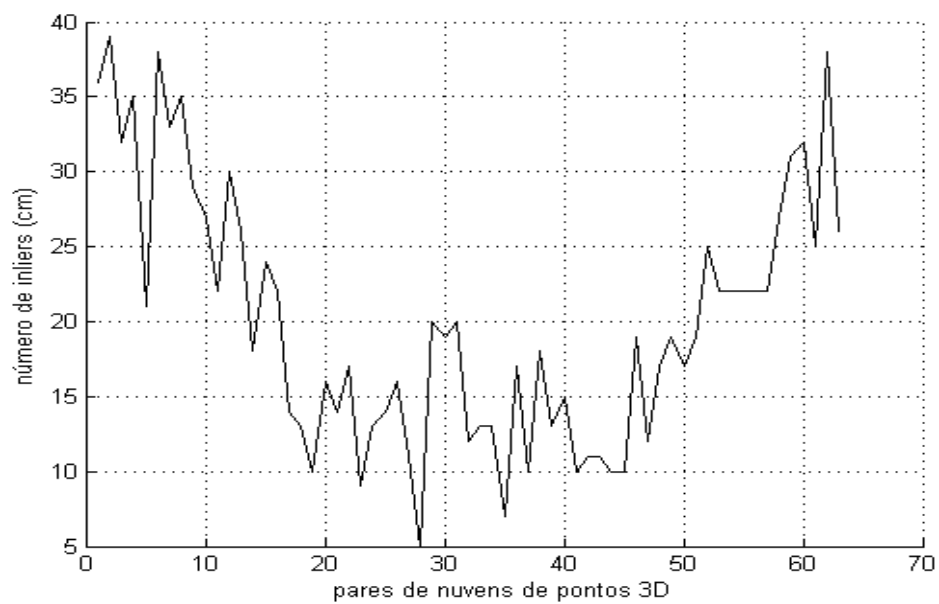
A REMQ dos experimentos referentes aos cenários F, G e H é apresentada na Tabela 3.

TABELA 3 - VALORES MÁXIMOS E MÍNIMOS DA REMQ PARA OS CENÁRIOS F, G e H

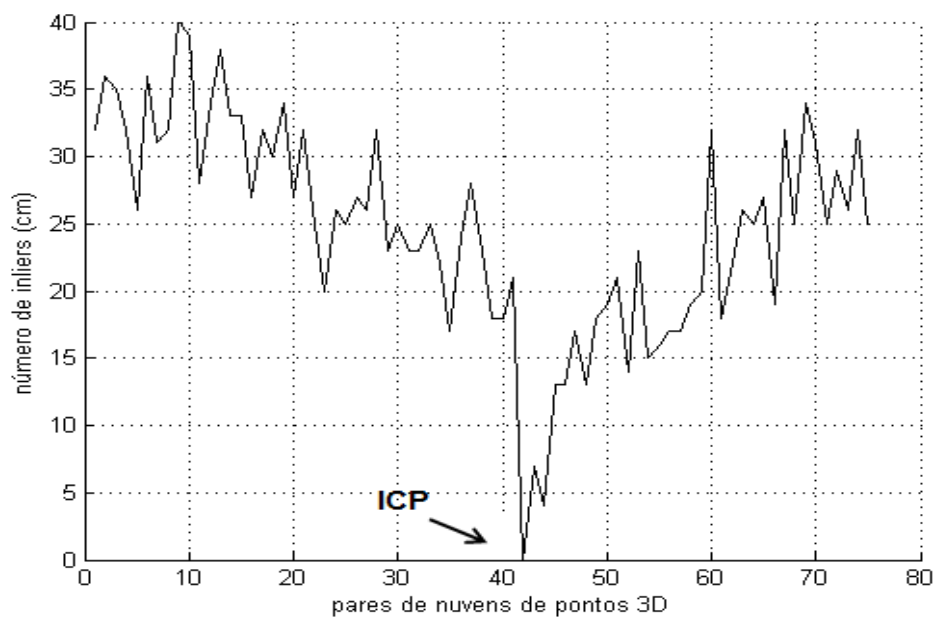
Experimento	REMQ (cm)
Cenário F	Min: 0,19 Max: 2,50
Cenário G	Min: 0,35 Max: 3,52
Cenário H	Min: 1,21 Max: 2,91

FONTE: O autor (2015)

Embora à distância mapeada seja curta, naqueles ambientes considerados pouco texturizados o método de registro grosseiro falha já que menos de 3 pontos podem ser detectados pelo SIFT ou até mesmo após a remoção de correspondências *outliers* restarem, portanto, menos de 3. A Figura 24 mostra um gráfico de *inliers* que foram encontrados no registro grosseiro dos pares de nuvens de pontos dos cenários I e J.



(a)

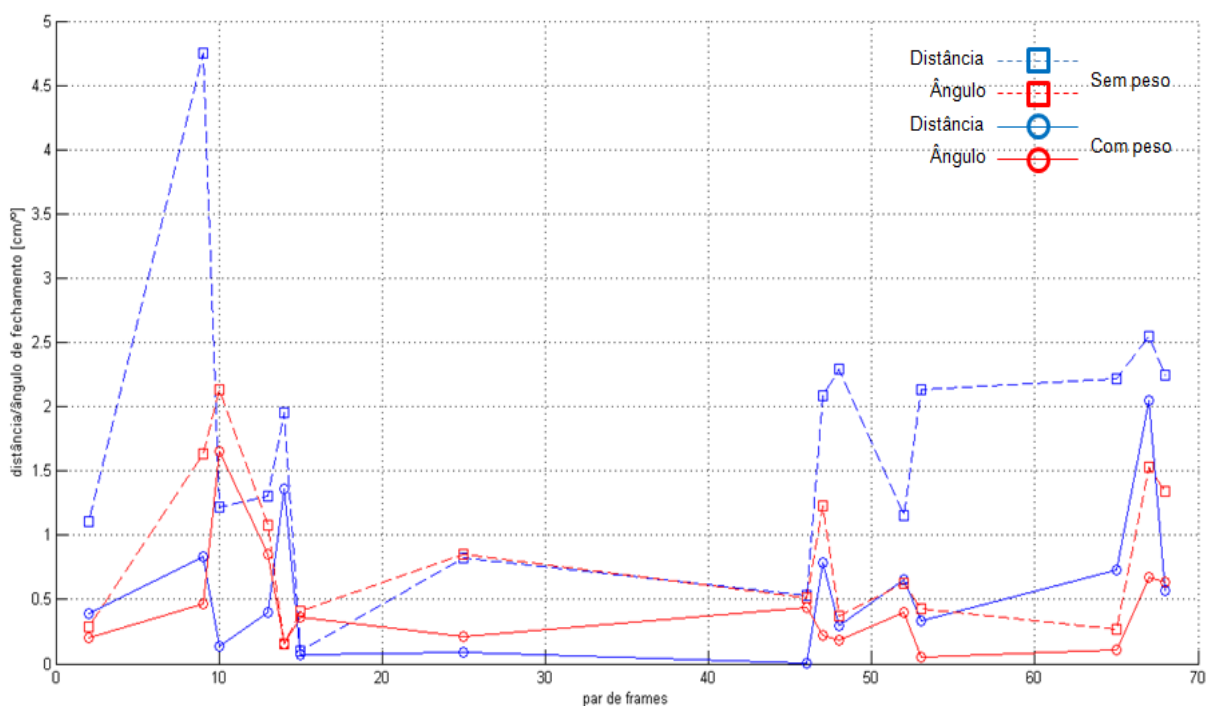


(b)

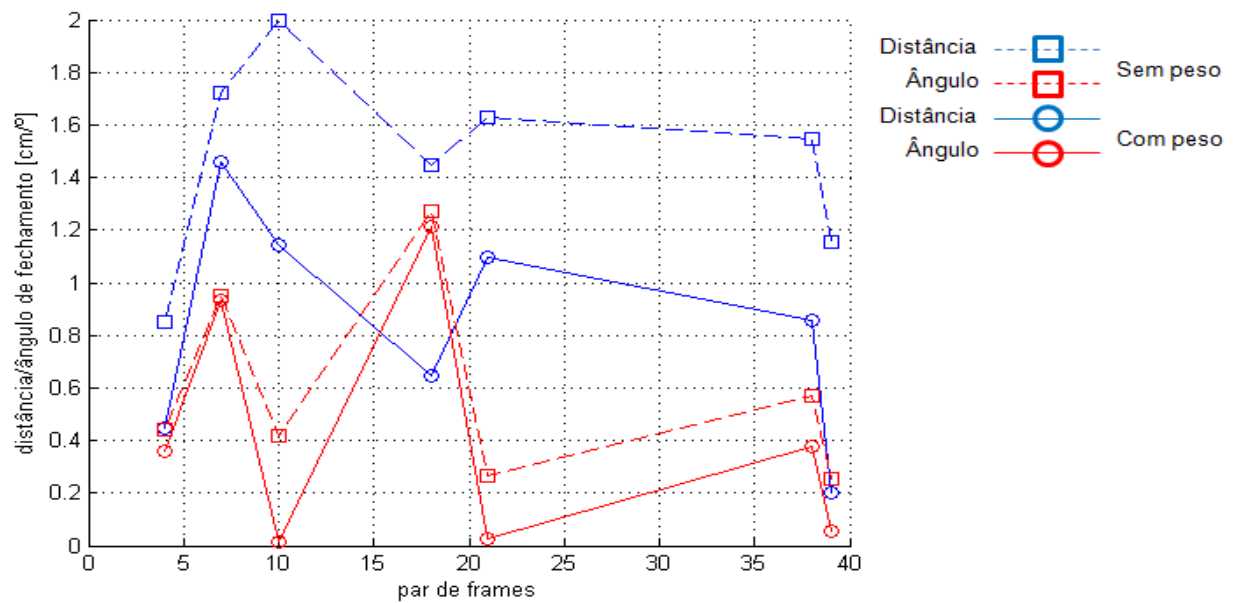
FIGURA 24 - *INLIERS* ADVINDOS DO REGISTRO GROSSEIRO COM O MÉTODO SIFT. (a) CENÁRIO I; (b) CENÁRIO J
 FONTE: O autor (2015)

O gráfico mostrado na Figura 24a revela que conforme o sensor percorre sua trajetória o número de *inliers* decai chegando a apenas 5 correspondências, e então, o número volta a subir assim que o *loop* volta, aproximadamente, ao seu ponto de origem. Diante situações como essa, o método sugerido neste trabalho se comporta bem como mostra a Figura 23b referente ao gráfico da REMQ das coordenadas desse cenário sendo que o maior valor não passou de 0,7 cm para a coordenada *X*, por exemplo. Já o gráfico mostrado na Figura 24b ilustra um caso extremo, como citado anteriormente, onde o SIFT falha e somente o algoritmo ICP é utilizado.

A Figura 25 contém dois gráficos que apresentam as distâncias e ângulos de fechamento para sequências de pares de nuvens de pontos sorteadas aleatoriamente dos cenários D e E após o registro dos pares com e sem pesos. As sequências foram ordenadas seguindo a sucessão dos pares. Essas distâncias e ângulos de fechamento representam as diferenças de translação e rotação (em módulo) de registros com e sem peso. Para chegar a esses valores, foram realizadas as diferenças entre os parâmetros de transformação (rotação e translação) obtidos no registro dos pares de nuvens de pontos 3D posterior com o antecessor.



(a)



(b)

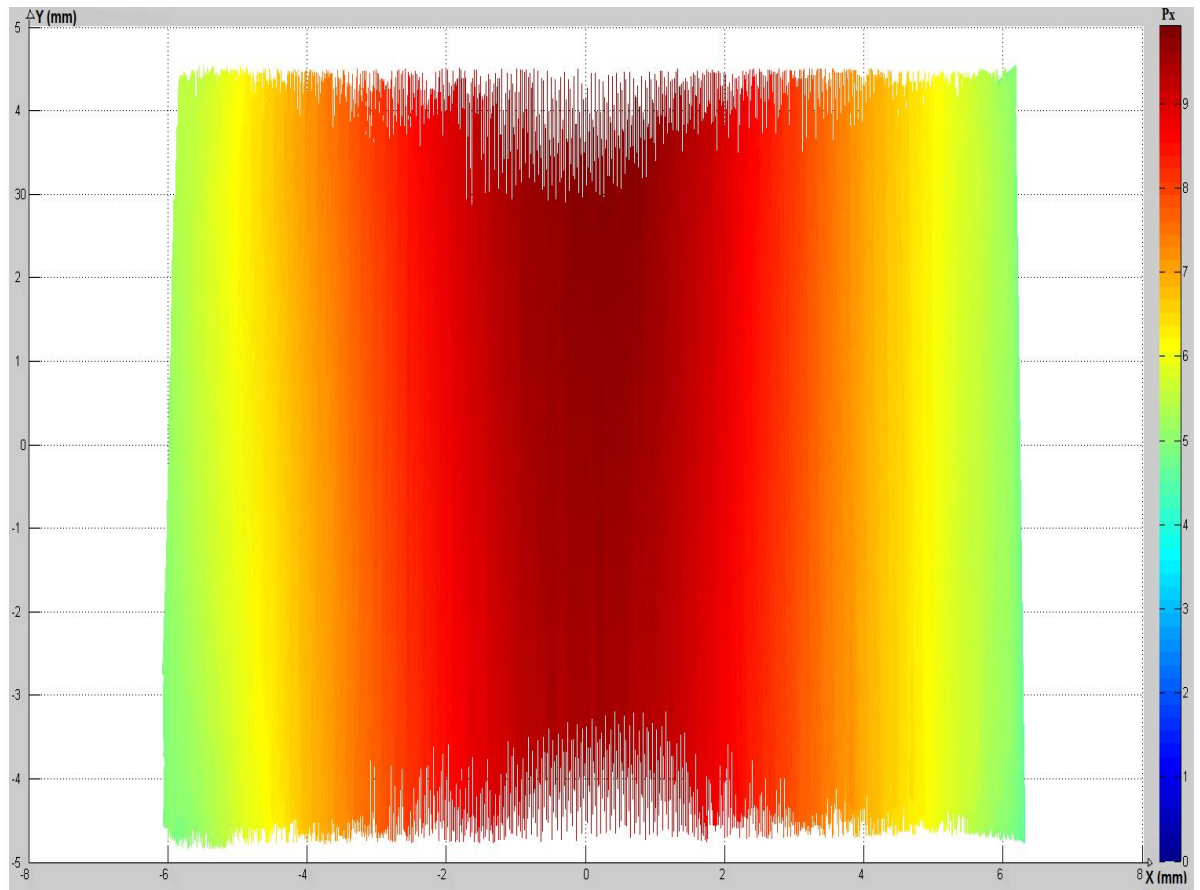
FIGURA 25 - DISTÂNCIA (AZUL) E ÂNGULO (VERMELHO) DE FECHAMENTO PARA SEQUÊNCIA DE REGISTROS ENTRE PARES DE NUUVENS DE PONTOS COM E SEM PESO. (a) CENÁRIO D; (b) CENÁRIO E;

FONTE: O autor (2015)

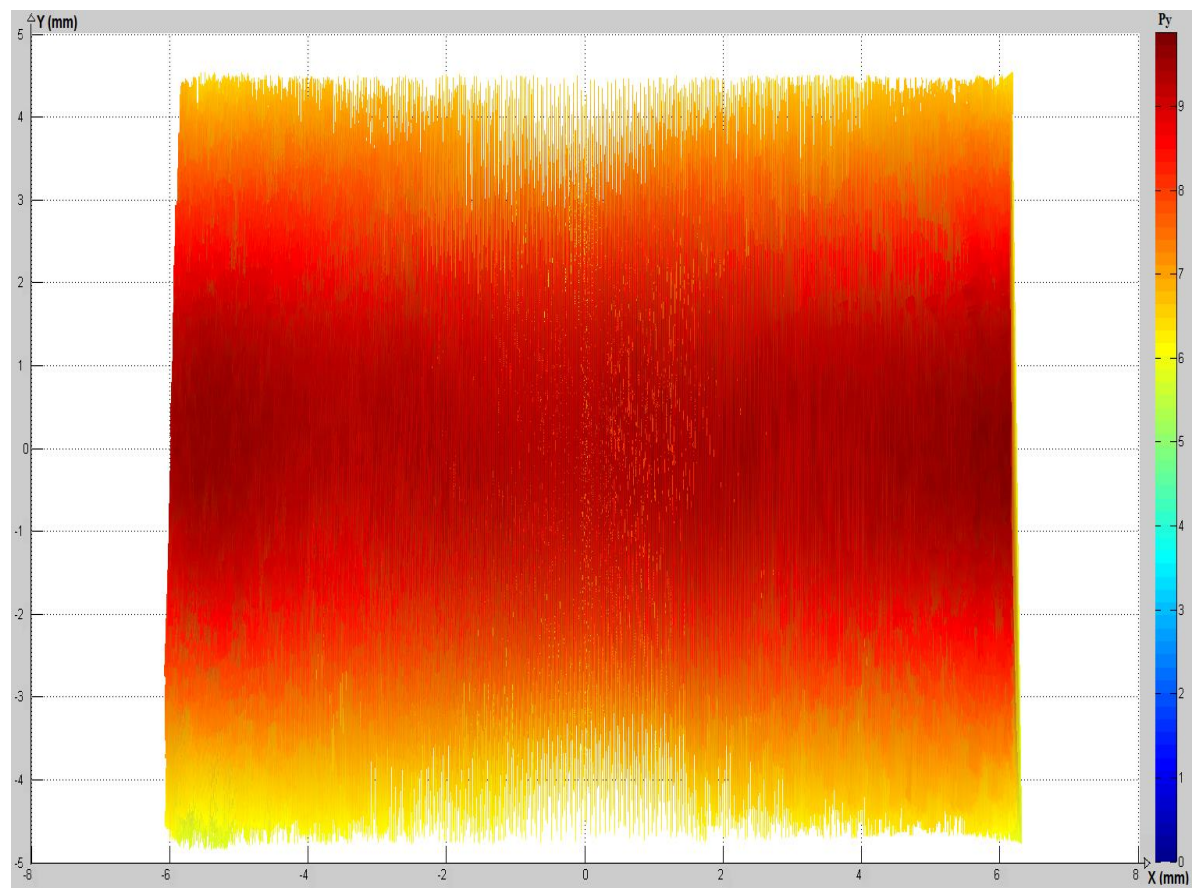
Como pode ser observado, ambas as distâncias e os ângulos de fechamento são melhoradas como resultado da utilização de pesos.

A explicação do não uso de pesos em XYZ reside em como a precisão teórica das medidas em X e Y são calculadas. Elas são feitas em função das coordenadas no espaço imagem, corrigidas dos erros sistemáticos, assim, os valores obtidos são piores à medida que o ponto se afasta do centro geométrico da imagem. Com a finalidade de mostrar esse efeito foi feito um experimento usando como referência uma nuvem de pontos 3D obtida pelo sensor Kinect em uma parede plana homogênea. A Figura 26 mostra a influência dos pesos baseados no erro randômico teórico das coordenadas X , Y e Z .

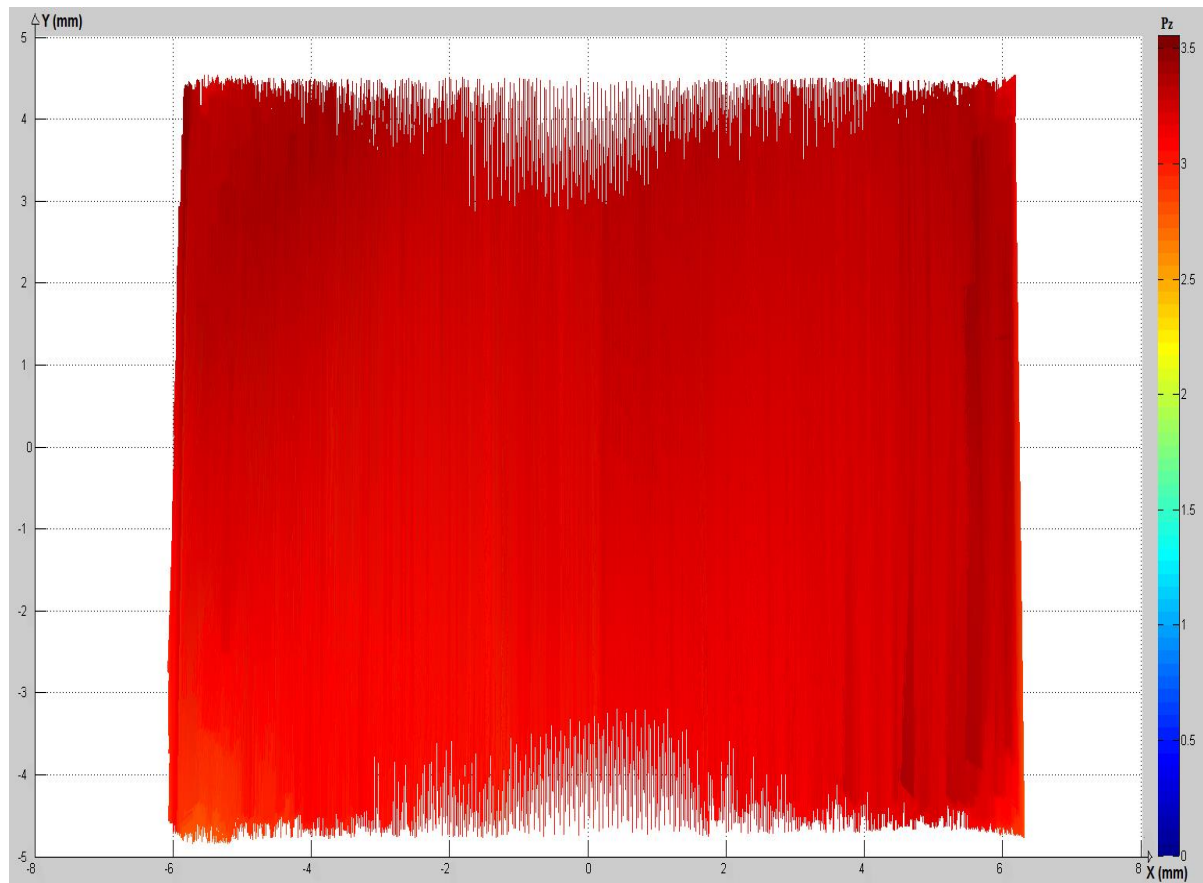
Nas Figuras 26a,b, conforme o ponto se afasta do centro geométrico da nuvem de pontos 3D menor é a contribuição dos pesos das observações nas coordenadas em X e Y (do azul ao vermelho e variando de 0 a próximo de 10 na escala de cores, sendo que quanto mais escura a cor maior a contribuição do peso). Esta redução na contribuição de pontos periféricos causa um efeito negativo no registro dos pares de nuvens de pontos. Já a Figura 26c revela a contribuição homogênea dos pesos em Z para qualquer região da nuvem de pontos 3D (também variando do azul ao vermelho e de 0 a aproximadamente 3,5).



(a)



(b)



(c)

FIGURA 26 - DISTRIBUIÇÃO DOS VALORES DE PESO PARA AS OBSERVAÇÕES. (a) INFLUÊNCIA DOS VALORES DE PESOS EM X; (b) INFLUÊNCIA DOS VALORES DE PESOS EM Y; (c) INFLUÊNCIA DOS VALORES DE PESOS EM Z.

FONTE: O autor (2015)

É inevitável destacar que no caso dos pesos em Z quanto mais próximo o objeto do sensor, melhor sua contribuição no processo de ajustamento de observações.

5. CONCLUSÕES E RECOMENDAÇÕES

Este trabalho apresentou um método adaptativo para o mapeamento 3D de ambientes internos com dados derivados de câmeras RGB-D. O método proposto utiliza tanto informações de profundidade quanto das imagens RGB do sensor. As principais características do método são: o uso de uma técnica robusta para a devida associação entre as imagens RGB e as nuvens de pontos 3D; a inclusão de pesos baseado no erro randômico teórico das medidas de profundidade; a subdivisão da nuvem de pontos de referência em células, que, além de organizar o processo de estabelecimento de correspondências, reduz o custo computacional do algoritmo ICP; e, o desenvolvimento de um modelo linear para ajustamento simultâneo dos pares de nuvens de pontos 3D tratando o problema de detecção de lugares revisitados em modo pós-processamento a partir da Matriz Variância Covariância dos parâmetros de transformação. O método de Zhang foi usado neste trabalho para a calibração dos sensores RGB e IR, bem como da orientação exterior entre os sensores. Como o foco principal deste trabalho não era a calibração, o método de Zhang foi empregado por ser totalmente automático e flexível para a nossa aplicação.

A detecção e correspondência automática nos pares de imagens RGB usando o algoritmo SIFT apresentou-se extremamente eficiente, pois praticamente não existe variações em rotação e mudança de ponto de vista. Quando é efetuada a associação entre os pontos SIFT 2D e seus correspondentes na nuvem de pontos 3D usando o método proposto neste trabalho, a correspondência é apurada, uma vez que a associação entre os pontos na imagem e na nuvem de pontos 3D depende das compensações angulares e lineares entre as imagens RGB e IR.

Dentro da etapa de refinamento dos parâmetros de transformação obtidos na etapa do registro grosseiro dos pares de nuvens de pontos 3D, o processo de correspondência ponto-a-ponto é feito de forma rápida e organizada, uma vez que apenas os centroides das células das nuvens de pontos de referência e seus correspondentes nas nuvens de pontos atuais são empregados. Além do mais, para esta etapa foi utilizada a técnica de busca do vizinho mais próximo *k-d tree*. Duas restrições geométricas foram propostas para diminuir o número de falsos positivos: a distância mínima e a variação angular das linhas retas que ligam os pontos

correspondentes adaptados ao valor da REMQ. O método proposto também apresenta uma técnica adaptativa para o critério de convergência baseado no inverso da distância do REMQ. Esta tarefa proporcionou uma redução significativa do tempo de processamento computacional. O método proposto apresentou uma melhoria significativa no registro do conjunto de nuvens de pontos 3D com ponderação na coordenada Z obtidas com o Kinect.

O método proposto para análise da consistência global é bastante simples, uma vez que emprega uma função linear e permite que as observações possam ser ajustadas simultaneamente, propiciando melhor precisão no ajustamento das observações, ao contrário dos métodos propostos na literatura que, geralmente, descartam as informações a medida que o erro de fechamento é detectado.

De forma geral, o método proposto apresenta como maior vantagem não necessitar de um número mínimo de *inliers* detectados pelo SIFT para determinar os parâmetros de transformação do registro, sendo que neste caso, o algoritmo ICP por si próprio deve ser capaz de realizar o registro das nuvens de pontos juntamente com o ajustamento global refinando esses parâmetros. Portanto, este método pode ser empregado em ambientes com regiões pouco texturizadas ou composta por muitos objetos com material especular ou de alta absorção de energia radiante na região do infravermelho próximo.

A critério de comparação, as nuvens de pontos dos cenários D, I e E foram obtidas a uma distância média de 2 metros, e 98,85% dos registros apresentaram uma acurácia relativa inferior a 11mm, que é aproximadamente a resolução da profundidade à essa distância, segundo Khoshelham e Elberink (2012). Em Endres *et al.* (2012), em um ensaio com 687 frames, as acurácias relativas encontradas foram de 4,2 cm e 2,5° para a translação e rotação, respectivamente, enquanto que com o método proposto, o maior valor encontrado foi de 3,52 cm para o cenário G. É importante frisar que o método se comportou de forma adequada, uma vez que era esperada uma redução nos valores de REMQ conforme a convergência do algoritmo, pressupondo que nenhuma tendência significativa influenciou o modelo matemático.

A detecção e remoção de *outliers* é uma tarefa complexa e exige maiores investigações, uma vez que eles afetam a estimativa dos parâmetros de transformação e, conseqüentemente, o cálculo do REMQ da média das discrepâncias. Para a calibração do sistema e dos sensores, em trabalhos futuros

pretendemos empregar modelos mais rigorosos, inclusive implementar um modelo que considere no processo de calibração o dispositivo LASER do Kinect. Também é recomendado, como trabalhos futuros, investigações sobre abordagens ICP ponto-a-plano no processo de registro dos pares de nuvens de pontos 3D, além da inclusão de novas restrições geométricas no processo de estabelecimento de correspondências e aplicar técnicas de filtragem robustas que reduzam o número de *outliers* presentes na nuvem de pontos 3D.

REFERÊNCIAS

AL-MANASIR, K.; FRASER, C. S. **Registration of terrestrial LASER scanner data using imagery**. The Photogrammetric Record, v. 21, p. 255-268, 2006.

ARUN K. S.; HUANG T. S.; BLOSTEIN S. D. **Least Square fitting of two 3-D point sets**. IEEE trans. Patt. Anal. Machine Intell. v. PAMI-9, n. 95, p. 698-700, 1987.

BACHRACH, A.; PRENTICE, S.; HE, R.; HENRY, P.; HUANG, A. S.; KRAININ, M.; MATURANA, D.; FOX, D.; ROY, N. **Estimation, planning, and mapping for autonomous flight using an RGB-D camera in GPS-denied environments**. The International Journal of Robotics Research, v. 31, p. 1320-1343, 2012.

BAE, K; LICHTI, D. D. **A Method for Automated registration of Unorganised Point Clouds**. ISPRS Journal of Photogrammetry and Remote Sensing, v. 63, p. 36-54, 2008

BARNEA, S.; FILIN, S. **Keypoint Based Autonomous Registration of Terrestrial LASER Point-Clouds**. ISPRS Journal of Photogrammetry and Remote Sensing, v. 63, p. 19-35, 2008.

BEINAT, A.; CROSILLA, F.; SEPIE, F. **Automatic morphological pre-alignment and global hybrid registration of LIDAR close range Images**. ISPRS Journal of Photogrammetry and Remote Sensing, Dresden – Germany, v. 36, parte 5, p 6, 2006. Disponível em: <http://www.isprs.org/proceedings/XXXVI/part5/paper/1243_Dresden06.pdf>. Acesso em: 3/11/2014.

BEIS, J.; LOWE, D. G. **Shape Indexing Using Approximate Nearest-Neighbour Search in High Dimensional Spaces**. In: Conference on Computer Vision and Pattern Recognition. Puerto Rico, p. 1000-1006, 1997.

BESL, P. J.; MCKAY, H. D. **A method for registration of 3-D shapes**. Pattern Analysis and Machine Intelligence, IEEE Transactions on, v. 14, p. 239-256, 1992.

BENDELS, G. H.; DEGENER, P.; WAHL, R., K, M.; KLEIN, R. **Image-based Registration of 3D-Range Data Using Feature Surface Elements**. In: 5ª International Conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage, Oudenaarde, Belgium, Eurographics Association, p. 115-124, 2004.

BLAIS G.; LEVINE M. D. **Registering multiview range data to create 3D computer objects**. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 17, n. 8, p. 820-824, 1995.

BRADSKI G.; KAEBLER A. **Learning OpenCv**. Editora O' Reilly, 2008.

BROWN, D. C. **Close-range camera calibration**. Photogramm, v. 37, p. 855-866 1971.

BROWN, M.; LOWE, D .G. **Invariant features from interest point groups**. In: British Machine Vision Conference. Cardiff, Wales, p. 656-665, 2002.

CHEN, Y.; MEDIONI, G. **Object Modelling by Registration of Multiple Range Images**. In: IEEE International Conference on Robotics and Automation Image, 1991, Sacramento CA. p. 2724-2729. Vision Comput., v. 10, p. 145-155, 1992. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=132043>>. Acesso em 24/06/2014.

CHOW J. C. K.; LICHTI, D. D. **Photogrammetric Bundle Adjustment with Self-Calibration of the PrimeSense 3D Camera Technology: Microsoft Kinect**. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 1, 2013.

CHOW J. C. K.; LITCH D. D.; HOL J. D., BELLUSCI; LUINGE H. **IMU and Multiple RGB-D Camera Fusion for Assisting Indoor Stop-and-Go 3D Terrestrial LASER Scanning. Robotics**. 3ª Edição, p. 247-280, 2014.

DALMOLIN, Q. **Ajustamento por Mínimos Quadrados**. 2ª edição, Curitiba-PR, Curso de Pós-Graduação em Ciências Geodésicas, v. 1, 175p, 2004.

DALMOLIN, Q.; SANTOS, D. R. **Ajustamento de Observações para Geotecnologias**. 1ª. Edição, Curitiba-PR, Curso de Pós-Graduação em Ciências Geodésicas, v. 1, 166p, 2014.

DOU, M.; GUAN, L.; FRAHM, J-M.; FUCHS, H. **Exploring High-Level Plane Primitives for Indoor 3D Reconstruction with a Hand-held RGB-D Camera**. In: Computer Vision - ACCV 2012 Workshops, Daejeon, Korea, p. 94-108, 2013.

DRYANOVSKI, I.; JARAMILLO, C.; JIZHONG, X. **Incremental registration of RGB-D images. Robotics and Automation (ICRA)**, IEEE International Conference on, p. 1685-1690, 2012.

DU, H.; HENRY, P.; REN, X.; CHENG, M.; GOLDMAN, D.B.; SEITZ, S.M.; FOX, D. **Interactive 3D modeling of Indoor Environments with a Consumer Depth Camera**. Proceedings of the 13^o International Conference on Ubiquitous Computing, Beijing, China, 2011.

EGGERT, D. W.; FITZGIBBON, A. W.; FISHER, R. B. **Simultaneous Registration of Multiple Range Views for Use in Reverse Engineering of CAD Models**. Computer Vision and Image Understanding, v. 69, ed. 3, p. 253-272, 1998. Disponível em: <http://homepages.inf.ed.ac.uk/rbf/MY_DAI_OLD_FTP/rp804.pdf>. Acesso em: 12/07/2014.

ELLEKILDE, L.-P.; HUANG, S.; VALLSMIRO, J.; DISSANAYAKE, G. **Dense 3D Map Construction for Indoor Search and Rescue**. In: Journal of Field robotics, v. 24, p. 71-89, 2007

ENDRES, F.; HESS, J.; ENGELHARD, N.; STURM, J.; CREMES, D.; BURGARD, W. **An evaluation of the RGB-D SLAM system, Robotics and Automation (ICRA)**, IEEE International Conference on, p. 1691-1696, 2012.

ENGELHARD, N.; ENDRES, F.; HESS, J.; STURM, J.; BURGARD, W. **Realtime 3D Visual SLAM with a Hand-Held RGB-D Camera**. RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden, 2001.

FAUGERAS O. D. HELBERT M., **The representation, recognition and location of 3-D objects**. Int. J. Robotic Res. v. 5, n.3, p. 27-52, 1986.

FENG, C. L.; HUNG, Y. S. **A Robust Method for Estimating the Fundamental Matrix**. In: VIIth Digital Image Computing: Techniques and Applications, Sydney, 10-12 Dec. 2003,

FIORAIO, N.; KONOLIGE, K. **Realtime visual and Point Cloud SLAM**. In: Velodynelidar, LIDAR, 2011. Disponível em: <http://www.velodynelidar.com/lidar/hdlpressroom/pdf/papers/journal_papers/Realtime%20Visual%20and%20Point%20Cloud%20SLAM.pdf>. Acesso em: 23/11/2014.

FISCHLER, M. A.; BOLLES, R. C. **Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography**. Communications of the ACM, v. 24, n. 6, p. 381-395, 1981.

FREEDMAN, B.; SHPUNT, A.; MACHLINE, M.; ARIELI, Y. **Depth mapping using projected patterns**. Prime Sense Ltd, United States, 2010.

FRIEDMAN, J. H.; BENTLEY, J. L.; FINKEL, R. A. **An algorithm for finding best matches in logarithmic expected time.** ACM Transaction on Mathematical Software, 3(3), p. 209–226, 1977.

GEMAEL, C. **Introdução ao ajustamento de observações: aplicações geodésicas.** Curitiba, UFPR, 1994.

GODIN, G.; RIOUX, M.; BARIBEAU, R.; **Three-dimensional registration using range and intensity information.** In: Videometrics III (SPIE), Boston, MA, v. 2350, p. 279-290, 1994. Disponível em: <<http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=981394&resultClick=1>>. Acesso em: 10/10/2014.

GOLUB, G.; van LOAN, C. **Matrix computation,** John Hopkins University Press, Baltimore, Maryland, 3rd, 1996.

GONZALES, R. C.; WOODS, R. E. **Processamento de Imagens Digitais.** Tradução de Roberto Marcondes César Junior e Luciano da Fontoura Costa – São Paulo, Editora Edgard Blucher Ltda, 2000.

GRANT, D.; BETHEL, J.; CRAWFORD, M. **Point-to-plane registration of terrestrial LASER scans.** ISPRS Journal of Photogrammetry and Remote Sensing 72(0), p. 16-26, 2012.

GREENSPAN M.; YURICK M. **Approximate K-D Tree Search for Efficient ICP.** In: 4^a International Conference on 3-D Digital Imaging and Modeling, IEEE Transactions on, 2003.

GRISSETTI, G.; STACHNISS, C.; GRZONKA, S.; BURGARD, W. **A Tree Parameterization for Efficiently Computing Maximum Likelihood Maps using Gradient Descent.** Proceedings of Robotics: Science and Systems. Atlanta, GA, USA, 2007.

GRUEN, A.; AKCA, D. **Least Square 3D Surface and Curve Matching.** ISPRS Journal of Photogrammetric and Remote Sensing, v 59, n. 3, p.151-174, 2005.

HALLERT, B. **Photogrammetry,** McGraw-Hill, New York (NY), 1960.

HARALICK, R. M.; JOO H.; LEE C.; ZHUANG X.; Vaidya G. V.; KIM M. B. **Pose estimation from corresponding point data**. In: Machine Vision for Inspection and Measurement, New York Academic, 84 p., 1989.

HARTLEY, R., e ZISSERMAN, R., **Multiple view and geometry in Computer Vision**, 2nd, Cambridge University Press, Cambridge, U.K., 655 p., 2003.

HENRY, P.; KRAININ, M.; HERBST, E.; REN, X.; FOX, D. **RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments**. In: Symposium on Experimental Robotics (ISER), 2010.

HENRY, P.; KRAININ, M.; HERBST, E.; REN, X.; FOX, D. RGB-D mapping: **Using Kinect-style depth cameras for dense 3D modeling of indoor environments**. The International Journal of Robotics Research, v. 31, p. 647-663, 2012.

HORN B. K. P. **Closed-form solution of absolute orientation using unit quaternions**, Journal Optical Society of America, v. p. 629-642, 1987. Disponível em: <http://people.csail.mit.edu/bkph/papers/Absolute_Orientation.pdf>. Acesso em 08/08/2014.

HUHLER, B.; JENKE, P.; STRASSER, W. **On On-the-Fly Scene Acquisition with a Handy Multisensor-System**, Int. J. Intell. Syst. Technol. Appl., v. 5, p. 255-263, 2008. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.330.3103&rep=rep1&type=pdf>>. Acesso em: 11/11/2014.

KHOSHELHAM, K.; ELBERINK, S. O. **Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications**: Sensors, v. 12, p. 1437-1454, 2012.

KHOSHELHAM, K., DOS SANTOS, D. R. e VOSSELMAN, G., **Generation and weighting of 3D point correspondences for improved registration of RGB-D data**. In: ISPRS Annals Volume II-5/W2: ISPRS Workshop LASER scanning, 11-13 November 2013, Antalya, Turkey / ed. by M. Scaioni et al Antalya: ISPRS, ISSN: 2194-9050. p. 127-132, 2013.

KÜMMERLE R. G.; GRISETTI H. S.; KONOLIGE B. **G²O: A General Framework for Graph Optimization**, IEEE International Conference on Robotics and Automation, Shanghai International Conference Center, Shanghai, China, 2011.

LINDBERG, T. Scale-space theory: A basic tool for analysing structures at different scales. In: Journal of Applied Statistics, 21, 2, p. 224-270, 1994.

LOURAKIS, M. I. A.; ARGYROS, A. A. **SBA: A Software Package for Generic Sparse Bundle Adjustment**. ACM Transactions on Mathematical Software (ACM), 36 (1), p. 1–30, 2009.

LOWE, D. G. **Object Recognition from Local Scale-Invariant Features**. In: International Conference on Computer Vision, Corfu, Set 1999. Disponível em: <<http://www.cs.ubc.ca/~lowe/papers/iccv99.pdf>>. Acesso em 10/10/2014.

LOWE, D. **Distinctive Image Features from Scale-Invariant Keypoints**. In: International Journal of Computer Vision, v. 60, p. 91-110, 2004.

LU, F.; MILIOS, E. **Globally Consistent Range Scan Alignment for Environment Mapping**. Netherlands, Autonomous robots, Kluwer Academic Publishers, Springer, p. 333-349, 1997. Disponível em: <<http://link.springer.com/article/10.1023%2FA%3A1008854305733>>. Acesso em: 14/12/2014.

MAGEE, M. J.; BOYTER, B.A.; CHIEN, C.-H.; and AGGARWAL, J.K.; **Experiments in Intensity Guided Range Sensing Recognition of Three-Dimensional Objects**. In: Pattern Analysis and Machine Intelligence, IEEE Transactions on, v. PAMI-7, p. 629-637, 1985.

MASUDA, T.; SAKAUE, K.; YOKOYA, N. **Registration and Integration of Multiple Range Images for 3-D Model Construction**. In: 13^a International Conference on Pattern Recognition, v. 1, p. 879-883, 1996.

MAY, S.; DROESCHEL, D.; HOLZ, D.; FUCHS, S.; MALIS, E.; HERTZBERG, J. **Three-dimensional mapping with time-of-flight cameras**. J. Field Robot, v. 26, p. 934-965, 2009.

MENNA, F.; REMONDINO, F.; BATTISTI, R., NOCERINO, E. **Geometric investigation of a gaming active device**. SPIE Videometrics Range Imaging and Applications XI, v. 8085, 2011.

MIKHAIL, E. M.; BETHEL, J. S.; MCGLONE J. C. **Introduction to modern photogrammetry**, Inc. New York: John Wiley & Sons, 479 p, 2001.

MOFFIT, F.; MIKHAIL, E. M. Photogrammetry, 3^a edition, Harper & Row, New York (NY), 1980.

MORE, J. **The levenberg-marquardt algorithm: Implementation and theory**. In: Numerical Analysis, Lecture Notes in Mathematics, Editora G.A. Watson, Springer-Verlag, 630 p, 1997.

MOSTOFI N.; MOUSSA A.; ELHABIBY M.; EL-SHEIMY N. **RGB-D Indoor Plane-based 3D-Modeling Using Autonomous Robot**. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Denver, Colorado – USA, ISPRS Technical Commission I Symposium, v. XL-1, p. 301-308, 2014.

NUCHTER, A.; LINGEMANN, K.; HERTZBERG, J. **Cached k-d tree Search for ICP Algorithms**. In: Proc. of the IEEE Conference 3-D Digital Imaging and Modeling, p. 419-426, 2007.

OLSON E.; LEONARD J.; TELLER S. **Fast iterative optimization of pose graphs with poor initial estimates**. In: Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), p. 2262–2269, 2006.

REISS, M. L. L. **Restituição Fotogramétrica Flexível de Imagens à Curta Distância: Estudo de Método e Desenvolvimento de Protótipo**. Dissertação (Mestrado em Ciências Cartográficas) – Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente, 161 p, 2002.

ROTH, G. **Registering two overlapping range images. 3-D Digital Imaging and Modeling**, In: Proc. of the Second International Conference on IEEE, Ottawa, Ont, p. 191-200, 1999. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=805349>>. Acesso em: 19/10/2014.

RUSINKIEWICZ, S.; LEVOY, M. **Efficient Variants of the ICP Algorithm**. In: Proc. of the 3^a International Conference On 3D Digital Imaging and Modeling – 3DIM, Tokyo, p.145-152, 2001. Disponível em: <http://www.cvl.iis.u-tokyo.ac.jp/~oishi/Papers/Alignment/Rusinkiewicz_Efficient_Variants_of_ICP.pdf>. Acesso em: 12/08/2014.

RUSU, R. B. **Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments**, Ph.D. thesis, Technische Universität München, 2009.

SANDE, C. V. D.; SOUDARISSANAME, S.; KHOSHELHAM, K.; **Assessment of Relative Accuracy of AHN-2 LASER Scanning Data Using Planar Features. Sensors** 10(9), p. 8198-8214, 2010. Disponível em: <<http://www.mdpi.com/1424-8220/10/9/8198/htm>>. Acesso em: 13/11/2014.

SEGAL, A.; HAEHNEL, D.; THRUN, S.; **Generalized-ICP**. In: Proc. of Robotics: Science and Systems (RSS), Seattle USA, 21 p, 2009. Disponível em: <<http://www.roboticsproceedings.org/rss05/p21.pdf>>. Acesso em: 26/08/2014.

SIMON, D. A. **Fast and Accurate Shape-Based Registration**, Ph. D. Dissertation, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1996.

STAMOS, I.; LEORDEANU, M. **Automated feature-based range registration of urban scenes of large scale**. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, v. 2, p. II-555-II-561, 2003.

STEINBRUCKER, F.; STURM, J.; CREMERS, D. **Real-time Visual Odometry from Dense RGB-D images**. Computer Vision Workshops (ICCV Workshops), IEEE International Conference on, p. 719-722, 2011.

THRUN, S.; MONTEMERLO, M. **The Graph SLAM Algorithm with Applications to Large-scale Mapping of Urban Structures**. The International Journal of Robotics Research, 403-429, 2006.

WEBER, T.; HANSCH, R.; HELLWICH, O. **Automatic registration of unordered point clouds acquired by Kinect sensors using an overlap heuristic**. In: ISPRS Journal of Photogrammetry and Remote Sensing, v. 102, p. 96–109, 2015.

WEIK, S. **Registration of 3-D partial surface models using luminance and depth information. 3-D Digital Imaging and Modeling**, In: International Conference on Recent Advances in IEEE, Ottawa, Ont, p. 93-100, 1997. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=603853>>. Acesso em: 06/09/2014.

WILLIAMS, J.; BENNAMOUN, M. **Simultaneous Registrations of Multiple Corresponding Point Sets**. Computer Vision and Image Understanding, 1ª ed. v. 81, p. 117-142, 2001.

WILLIAMS, V.A., BRAZIER, H.H.: **The method of adjustment of independent models. Huddersfield test strip**. The Photogrammetric Record (26), 123–130, 1965.

WORLD HEALTH ORGANIZATION, Urban population growth, Global Health Observatory (GHO), [Online]. Disponível: <http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en>. Acesso em 10/07/2015.

ZHANG, Z. **Iterative point matching for registration of free-form curves and surfaces**. Boston, International Journal of Computer Vision, v. 13, p. 119-152, 1994. Disponível em: <<http://research.microsoft.com/en-us/um/people/zhang/Papers/IJCV-94-ICP.pdf>>. Acesso em 06/07/2014.

ZHANG, Z. **Determining the Epipolar Geometry and its Uncertainty: A Review**. In: International Journal of Computer Vision, Kluwer Academic Publishers, Boston, 27(2), p. 161–198, 1998.

ZHANG, Z. **A Flexible New Technique for Camera Calibration**. Redmond (WA): Microsoft Research, p. 21, 2000. Relatório Técnico. Disponível em: <<http://research.microsoft.com/~zhang>>. Acesso em: 01/10/2014.